A Hybrid CNN–BiLSTM–Attention Deep Learning Framework for Sediment Transport in an Agricultural Gully System

Naziru Halilu[1,2,3,4]*

[1] Department of Agricultural and Bio Resources Engineering, Faculty of Engineering, Ahmadu Bello University, Zaria 810107, Nigeria

[2] Higher Technical School of Agricultural Engineering and Bioscience, Public University of Navarre, Pamplona 31006, Spain

[3] University of Trás os Montes and Alto Douro, Vila Real 5000 801, Portugal

[4] Agricultural University of Athens, 118 55 Athens, Greece

***Correspondence Author**

DOI: https://doi.org/10.51244/IJRSI.2026.1307000217

Received: 23 July 2026; Accepted: 28 July 2026; Published: 07 August 2026

ABSTRACT

Gully erosion threatens soil and water resources across agricultural landscapes, yet most predictive tools for small watercourses rely on simplistic empirical or linear models. This study develops a hybrid Conv1D-BiLSTM-Attention deep learning architecture to predict sediment transport rate in a gully system actively rehabilitated with Morning Glory (Ipomoea carnea) vegetative cover. A six-variable dataset (flow depth, channel slope, soil shear stress, flow velocity, channel breadth and average depth; 100 observations across five field scenarios) was compiled and linked to governing sediment-transport equations (continuity, Straub shear stress, Duboys and Meyer-Peter-Müller bed-load relations) and the Wu-Waldron root-reinforcement model. The hybrid network, trained with Adam under a weighted Huber loss with L2 regularisation and dropout using 10x-repeated 5-fold cross-validation, achieved a hold-out R^2 of 0.90 (RMSE = 0.45 kg/s/m, MAPE = 20.1%; bootstrap 95% CI: 0.72-0.97) and a cross-validated R^2 of 0.68 ± 0.36, outperforming a benchmark multilayer perceptron (R^2 = 0.62 ± 0.37) though not significantly different from either the MLP or multivariate linear regression (R^2 = 0.78 ± 0.28; Mann-Whitney U, p = 0.06 and 0.08). Permutation importance and Kernel SHAP identified flow depth and soil shear stress as dominant predictors, while Monte Carlo dropout produced 95% prediction intervals with 73% empirical coverage, somewhat under the nominal target at this sample size. Morning Glory reduced mean sediment transport by 25.5% at 1.0 m ponding depth but only 1.4% at 1.5 m, indicating diminishing vegetative control once hydraulic loading exceeds root-reinforcing capacity. This integrated, physically grounded and uncertainty-aware framework provides a reproducible basis for scaling hybrid deep-learning sediment transport prediction to larger agricultural landscapes.

Keywords: Hybrid deep learning; CNN-BiLSTM-Attention; Sediment transport; Gully erosion

INTRODUCTION

Background

Gully erosion is one of the most visible and damaging forms of land degradation in agricultural landscapes, reshaping channels, removing productive topsoil and delivering large sediment loads to downstream water bodies (Poesen et al., 2003). Sediment transport rate in agricultural watercourses is governed by an interacting set of hydraulic and geomorphic variables, flow depth, channel slope, soil shear strength, flow velocity and channel geometry, whose combined effect is rarely a simple linear function of any one variable (Kirkby and Bracken, 2009). In Nigeria, gully erosion is one of the most severe forms of land degradation, with thousands of active sites recorded across the south-eastern states and large cumulative losses of arable land and infrastructure (Igwe and Egbueri, 2018; Francisca et al., 2023), part of a global loss of tens of billions of tonnes of topsoil each year that threatens food security (Pimentel, 2006; Lal, 2001). Reliable prediction of sediment transport is central to planning erosion-control measures and evaluating interventions such as vegetative cover, and gully erosion has proved persistently difficult to arrest once initiated (Valentin et al., 2005; Morgan, 2005).

Traditional approaches rely heavily on empirical or physically based bed-load equations, most notably the Duboys (1879) shear-excess formulation and simple velocity-sediment regressions (Van Rijn, 1984a,b). Such models assume a fixed, low-order functional form and oversimplify the non-linear, multi-variable character of sediment transport, particularly where channel conditions change substantially, for example after an erosion-control measure is installed. A companion field study at the same watercourse used a single-variable linear regression between velocity and sediment transport rate to evaluate Morning Glory as a biological control; while it captured the broad trend, its single-variable formulation could not exploit the full information contained in the field measurements of depth, slope, soil shear and channel geometry.

Vegetative (Bioengineering) Approaches to Gully Erosion Control

Vegetative, or soil-bioengineering, measures are widely used to complement or replace structural erosion-control works, exploiting the mechanical and hydrological effects of plant roots and canopy cover (Gray and Sotir, 1996). Roots reinforce soil by increasing its effective shear strength through fibre-like reinforcement (Wu et al., 1979; Waldron, 1977; Wu, 2013) and by altering near-surface hydrology and aggregate stability, depending on root architecture and density (Ola et al., 2015). Species selected for gully rehabilitation are typically fast-growing and capable of developing a dense fibrous root system within a single growing season; vetiver grass (Chrysopogon zizanioides) is the most extensively documented example, shown to reduce erosion by up to 90% on established slopes (Truong and Loch, 2004; Truong et al., 2008).

Morning Glory (Ipomoea carnea Jacq.), locally known as Ana Kwari, is a fast-growing, drought-resistant perennial shrub with a fibrous, laterally spreading root system and a climbing habit (Kamal et al., 2017), widely naturalised along watercourses owing to its plastic root/shoot architecture and rapid propagation from cuttings (Shah et al., 2022). These traits motivated its selection as the vegetative control measure in this study, planted as stem cuttings in April 2024 using a double-density striking technique and monitored through germination, growth, canopy coverage, root development and effectiveness, as detailed below.

Deep Learning in Sediment Transport Prediction

Machine learning and deep learning methods have substantially improved sediment transport prediction by learning non-linear relationships directly from data. Deep neural networks have outperformed conventional regression for predicting suspended and total sediment load (Shakya et al., 2023), and feature-selection/deep-network frameworks have improved sediment discharge estimation in river basins (Kaloop et al., 2025). LSTM networks have outperformed conventional neural networks and support vector machines for sediment load prediction (Latif et al., 2023), and reviews consistently note deep learning's growing role relative to purely empirical or statistical models (Andualem et al., 2023). Earlier machine learning approaches (artificial neural networks, model trees) also demonstrated gains over regression-based rating curves once sufficient training data were available (Bhattacharya et al., 2007), and deep recurrent architectures are now common in rainfall-runoff and streamflow forecasting (Kratzert et al., 2018), building on an established tradition of neural network applications in hydrology (ASCE Task Committee, 2000; Solomatine and Ostfeld, 2008). Random forests and permutation-importance techniques have similarly been used to rank the factors controlling gully erosion susceptibility at landscape scale (Rahmati et al., 2017; Amiri et al., 2019), providing methodological precedent for the interpretability analysis in this study (Fisher et al., 2019; Breiman, 2001).

Despite this progress at river-basin and landscape scale, deep learning has rarely been applied to small, intensively monitored agricultural gully systems with modest sample sizes but detailed, scenario-based measurements. Within the deep learning literature, hybrid architectures combining convolutional feature extraction with recurrent or attention-based sequence encoding (e.g., CNN-LSTM, CNN-BiLSTM-attention) have increasingly outperformed single-mechanism architectures for hydrological and streamflow prediction by capturing local cross-variable patterns and longer-range dependencies simultaneously (Ng et al., 2023; Huang and Zhang, 2024), and attention mechanisms improve both accuracy and interpretability by learning which inputs most strongly influence a given prediction (Huang and Zhang, 2024). This study addresses the small-sample agricultural gully gap directly, applying such a hybrid convolutional-recurrent-attention architecture to a compact field dataset while situating the model within the governing physical equations of sediment transport and root-reinforced soil, and within the calculus of its own optimisation procedure.

Objectives of the Study

Using field data from a gully system within the Ahmadu Bello University Dam watercourse, this paper develops a hybrid CNN-BiLSTM-Attention deep learning architecture to predict sediment transport rate from flow depth, channel slope, soil shear stress, flow velocity, channel breadth and channel average depth. The objectives are to: (i) characterise the study watershed and channel network geospatially and set out the governing physical equations of sediment transport, including their integral-calculus formulation, and relate them to the field-measured variables; (ii) document the establishment and monitoring of Morning Glory as a vegetative control measure and interpret its effectiveness through the Wu-Waldron root-reinforcement model; (iii) compile a multi-scenario field dataset spanning pre- and post-control conditions merged with an independent channel geometric survey; (iv) design, train and validate a hybrid convolutional-recurrent-attention deep learning model, with a full mathematical formulation, calculus-based training procedure, and interpretability (permutation importance, Kernel SHAP) and uncertainty-quantification (Monte Carlo dropout) analyses; and (v) benchmark the hybrid model against a previously developed feedforward MLP and multivariate linear regression on the same six variables, to assess honestly the added value, and current limitations, of a more architecturally sophisticated approach at this data scale.

MATERIALS AND METHODS

Study Area and Data Source

The study site is located within Nigeria, in Kaduna State in the country's North-West geopolitical zone, specifically in Zaria Local Government Area on the campus watercourse of Ahmadu Bello University (ABU). At the national scale, Nigeria spans the transition from the Guinea Savanna of the Middle Belt to the drier Sudan and Sahel zones further north; Kaduna State sits within this Guinea Savanna belt and is drained by tributaries of the Kaduna River system. Zaria town, in southern Kaduna State, hosts the ABU main campus and its dam-fed watercourse, the specific agricultural gully system studied here. The surveyed reach falls within approximately 3.42–3.44°E and 7.22°–7.25°N (Figure 1), consistent with the DEM extent shown in Figure 1A.The study was carried out on a gully system within the watercourse of the Ahmadu Bello University (ABU) Dam, Zaria, Kaduna State, Nigeria, an agricultural landscape in the Nigerian Guinea Savanna. The site has been subject to active gully erosion and was rehabilitated using Morning Glory together with controlled ponding of water outside the gully as an experimental erosion-control strategy. Field data used to develop the model were drawn from a field campaign that measured channel and flow variables before and after the vegetative control measure, at two ponding depths, together with an independent geometric channel survey.

To place the surveyed 125 m channel reach in its landscape context, a geospatial characterisation of the contributing watershed was carried out from a 30 m SRTM digital elevation model (DEM), NDVI/EVI composites for 2014 and 2024, and the field survey data (Figure 1). The DEM (1A) shows the watershed rising from ~436 m in the south-west to 512 m along the north-eastern divide; the slope-gradient surface (1B) identifies the steepest terrain (>20%) along the upper reaches, coincident with the drainage-network/stream-order map (1D, Strahler order 1st-5th) and flow-accumulation surface (1C, maximum 14,820 cells). Overlaying these with vegetation indices identifies erosion source areas and susceptibility hotspots on the steep, sparsely vegetated upper slopes (1E-F), while low-gradient downstream reaches are the dominant deposition zones (1G). The 20 fixed-interval field sampling points sit within the erosion-susceptible zone (1H), corroborating the reach-scale governing-equation analysis presented below.

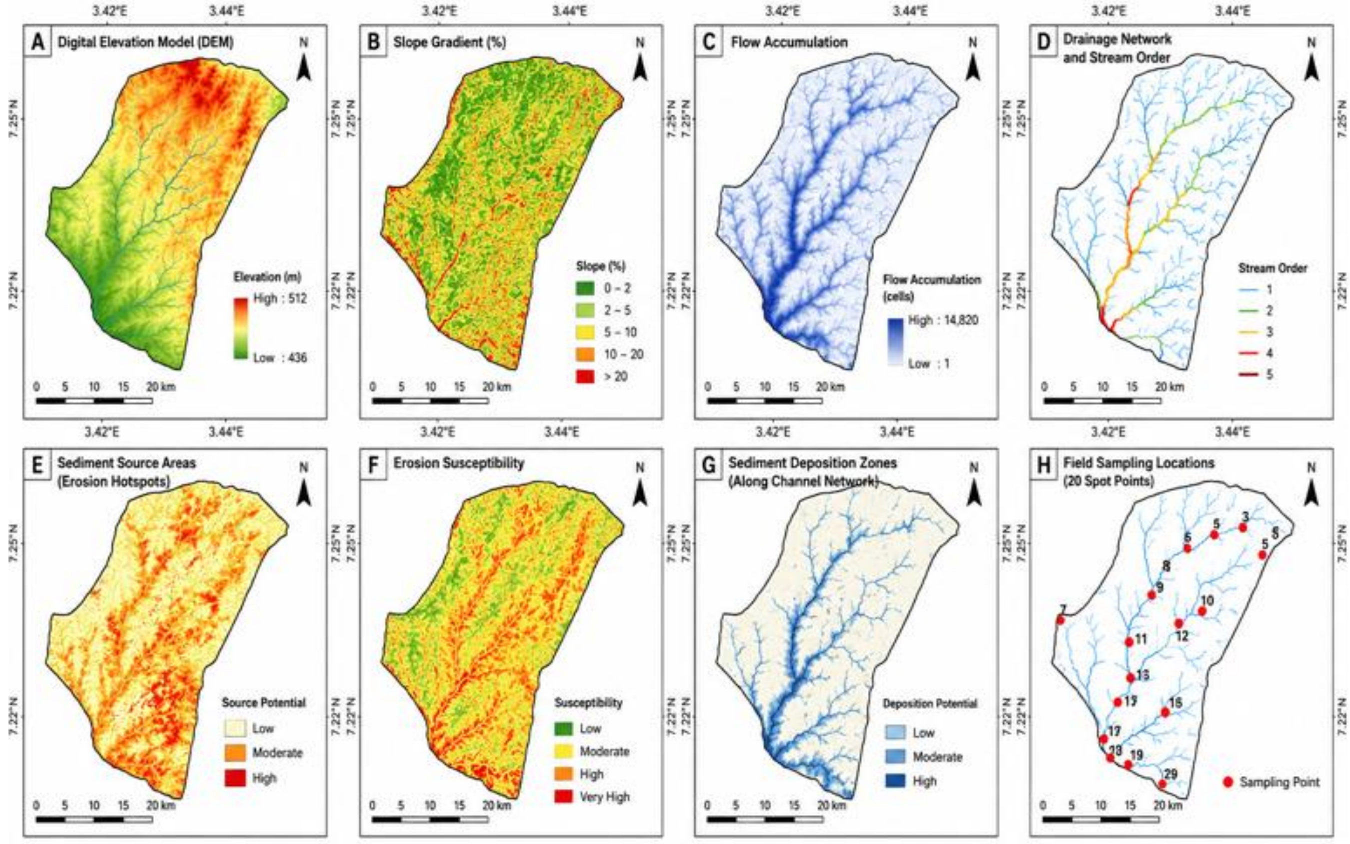

Figure 1. Geospatial characterisation of the study watershed: (A) digital elevation model; (B) slope gradient; (C) flow accumulation; (D) drainage network and Strahler stream order; (E) sediment source area; (F) erosion susceptibility; (G) sediment deposition zones; (H) field sampling locations (20 spot points).

Governing Equations of Sediment Transport

The field variables are linked by established relationships from open-channel hydraulics and sediment transport theory, motivating the choice of model inputs and providing a physical benchmark for interpreting the data-driven model.

The continuity equation relates discharge Q to flow velocity V, channel breadth B and flow depth D:

$$Q = V \times B \times D \qquad (1)$$

This motivates including channel breadth alongside flow depth and velocity, since for a given discharge, breadth and depth jointly determine velocity and hence erosive energy.

The boundary (tractive) shear stress exerted by flowing water, following Straub (1935), is:

$$\tau = \gamma DS \qquad (2)$$

where τ is boundary shear stress, γ the specific weight of water, D flow depth and S channel slope the basis for the soil shear stress variable measured in the field.

Duboys (1879) proposed that bed-load moves in stacked layers once applied shear exceeds a critical threshold:

$$q = k(\tau - \tau_c) \qquad (3)$$

where q is sediment transport rate, k the Duboys coefficient and τc the critical shear stress (both read from the Duboys graph). This equation was used in the field campaign to compute the sediment transport rate values forming the model's target variable.

For reference, the Meyer-Peter and Müller (1948) bed-load formula expresses transport as a power function of excess shear:

$$q_m = C(\tau - \tau_c)^{3/2} \qquad (4)$$

together with the Shields (1936) and Schoklitsch (1934) incipient-motion criteria, this corroborates the shear-excess principle of Equation (3).

Formulation of Sediment Yield

Because sediment transport rate q varies continuously with time and along the channel, the physically meaningful quantities for erosion-control planning, cumulative yield and cross-sectional discharge, are obtained by integrating q. Cumulative sediment yield over a monitoring period of duration T is:

$$Y = \int_0^T q\,(t)\,dt \qquad (5)$$

and, because q is expressed per unit channel width, total sediment discharge Qs across a cross-section of breadth B is:

$$Q_s = \int_0^B q\,(y)\,dy \approx \bar{q} \times B \qquad (6)$$

where $\bar{q}$ is the breadth-averaged transport rate. Equations (5)-(6) connect the pointwise model predictions to the time- and width-integrated sediment budgets needed for erosion-control design. Together, Equations (1)-(6) justify the six-variable predictor set used throughout this study.

Advanced Sediment Transport and Gully-Trapping Equations

Beyond the Duboys and Meyer-Peter-Müller shear-excess relations (Equations 3-4), several further established formulations describe bed- and suspended-load transport and, specifically, the trapping of sediment once it enters a gully or check-dam structure. These are included to situate the six-variable predictor set within the broader sediment-transport literature and to provide the physical basis for the trapping-efficiency analysis in Figure 13.

The Einstein (1950) bed-load function expresses transport through a dimensionless intensity parameter ψ and a dimensionless transport parameter Φ:

$$\psi = \frac{(\rho_s - \rho)}{\rho}\frac{d}{RS} \qquad (7)$$

$$\Phi = \frac{q_b}{\rho_s[g(\rho_s/\rho - 1)d^3]^{1/2}} \qquad (8)$$

where ρ_s and ρ are sediment and water density, d the characteristic grain size, R the hydraulic radius and q_β the bed-load transport rate per unit width; ψ and Φ are related through Einstein's empirical transport curve.

The Engelund and Hansen (1967) total-load formula, widely used for sand-bed channels, gives the dimensionless total sediment transport as a function of the Shields parameter θ:

$$0.05 = \frac{f}{2}\frac{q_t}{[(\rho_s/\rho - 1)gd^3]^{1/2}}\theta^{-5/2} \qquad (9)$$

with $\theta = \tau/[(\rho_s - \rho)gd]$ and f the friction factor, q_t the total sediment discharge per unit width.

The Ackers and White (1973) method expresses transport through a dimensionless sediment size Dgr and mobility number Fgr:

$$D_{gr} = d\left[\frac{g(\rho_s/\rho - 1)}{\nu^2}\right]^{1/3} \quad ; \quad G_{gr} = C\left(\frac{F_{gr}}{A} - 1\right)^m \qquad (10)$$

where ν is kinematic viscosity and A, C, m are grain-size-dependent coefficients.

Suspended-sediment concentration is described by the Rouse (1937) profile, relating concentration c at height y above the bed to a reference concentration c_a at height a:

$$\frac{c}{c_a} = \left[\frac{(D-y)}{y}\frac{a}{(D-a)}\right]^Z \qquad (11)$$

with Rouse number Z = $w_s/(\kappa u^*)$, w_s the sediment fall velocity, κ the von Kármán constant and u* the shear velocity; the van Rijn (1984a,b) reference concentration is:

$$c_a = 0.015\frac{d_{50}}{a}\frac{(T^{1.5})}{D_*^{0.3}} \qquad (12)$$

where T is the transport-stage parameter and D* the dimensionless particle parameter.

Bed-elevation change over time, and hence the volume of sediment retained or scoured within a gully reach, is governed by the Exner (1925) sediment-continuity equation:

$$(1-\lambda_p)\frac{\partial \eta}{\partial t} = -\frac{\partial q_s}{\partial x} \qquad (13)$$

where η is bed elevation, λ_p sediment porosity, x the along-channel coordinate and q_s the volumetric sediment transport rate; integrating Equation (13) along the 125 m surveyed reach (Table 1) over a monitoring interval gives the net deposition or erosion volume behind any vegetative or structural control.

An alternative, stream-power-based formulation (Bagnold, 1966) expresses the sediment transport capacity as:

$$q_s \propto \frac{\omega - \omega_c}{\tan\varphi} \qquad (14)$$

where ω = τV is the stream power per unit bed area, ωc a critical stream power and φ the submerged friction angle of the bed material.

Finally, the proportion of incoming sediment retained within a gully or behind a check-dam/vegetative barrier, the trap efficiency TE, is commonly estimated from the empirical Brune (1953) curve as a function of the capacity-to-inflow ratio C/I:

$$TE = 1 - \frac{1}{1 + 0.1(C/I)} \qquad (15)$$

or, for small structures such as the vegetated barrier used here, from Churchill's (1948) trapping-efficiency index η_c (sedimentation index, a function of flow-through velocity and retention period):

$$TE(\%) = 100 - \frac{100}{1 + 0.0021\,\eta_c} \qquad (16)$$

Equation (16) is applied directly below to express the observed pre-/post-control transport reduction (qsb versus qsa) as a trap efficiency, and underlies the trapping-dynamics panels of Figure 13. The corresponding sediment delivery ratio (SDR), linking gross upslope erosion Eg to the net sediment yield Ys actually delivered past the gully outlet, is:

$$SDR = \frac{Y_s}{E_g} = 1 - TE \qquad (17)$$

so that a higher trap efficiency (larger vegetative or structural retention) corresponds directly to a lower sediment delivery ratio at the watershed outlet.

Field Data Collection

Flow depth was measured with a stadia rod, channel slope with a Garmin GPS unit and clinometer/altimeter, and flow velocity by the floating method. Soil shear stress was estimated from Equation (2), and sediment transport rate from Equation (3) using the Duboys graph. Measurements were collected under five scenarios: (i) pre-control; (ii) pre-control at 1.0 m ponding; (iii) post-control at 1.0 m ponding; (iv) pre-control at 1.5 m ponding; and (v) post-control at 1.5 m ponding, with 20 paired observations per scenario (100 total, listed in Appendix A).

Channel Geometric Survey

An independent geometric survey covered the 125 m reach at 20 fixed spot intervals, recording channel breadth and two depth readings (from which average channel depth was computed) together with local bed slope. Matched slope values confirmed these 20 survey spots correspond exactly to the 20 hydraulic sampling points (maximum discrepancy $< 2\times10^{-3}$), allowing breadth and average depth to be merged with the hydraulic dataset as two scenario-invariant predictors, consistent with Equation (1). Table 1 reproduces the survey; channel depth peaked at spot 18, attributed to poor bed grading there.

Table 1. Channel geometric characteristics along the 125 m surveyed reach (20 spot intervals)

Spot	Breadth (m)	Depth 1 (m)	Depth 2 (m)	Average depth (m)	Slope (%)
1	1.78	2.10	2.20	2.15	0.90
2	1.95	2.20	2.01	2.10	0.67
3	1.92	1.83	1.85	1.84	1.20
4	2.01	1.95	1.90	1.92	0.95
5	2.20	2.80	2.86	2.83	0.90
6	2.30	2.50	2.54	2.52	0.70
7	1.94	3.00	2.54	2.77	1.10
8	2.70	2.20	2.23	2.22	0.76
9	2.30	2.50	2.23	2.37	0.56
10	2.20	2.30	2.31	2.31	2.10
11	1.90	1.95	1.95	1.95	1.30
12	3.00	1.67	1.65	1.66	1.40

13	2.20	1.53	1.53	1.53	0.96
14	1.98	1.65	1.96	1.96	0.96
15	2.01	1.83	1.84	1.84	1.05
16	2.23	1.76	1.67	1.72	1.12
17	2.52	1.10	1.10	1.10	1.10
18	3.35	1.23	1.20	1.22	1.51
19	2.96	1.76	1.70	1.73	1.32
20	2.81	2.01	2.01	2.01	1.08

Vegetative Erosion Control: Morning Glory (Ipomoea carnea) Establishment

Morning Glory was planted in early April 2024 from stem cuttings using a double-density striking technique to maximise early coverage and root development. The plant fully regenerated within ~1.5 months, attributed to its fibrous root structure, deep rooting habit and large leaf area index (Lawrence, 1983). Growth was monitored through six sequential stages (Table 2): planting, germination monitoring, growth monitoring, coverage evaluation, root-development evaluation, and an effectiveness analysis comparing pre- and post-control sediment transport.

Table 2. Monitoring stages and indicative timeline for Morning Glory establishment

Stage	Activity	Indicative timing
1. Planting	Insertion of stem cuttings at identified erosion-prone spots using double-density striking	Early April 2024 (day 0)
2.Germination monitoring	Observation of stem sprouting and initial rooting	Weeks 1-3
3.Growth monitoring	Checks for disease, pest infestation or stunted growth	Weeks 2-6
4.Coverage evaluation	Measurement of canopy ground coverage across planted spots	Weeks 3-6
5. Root development evaluation	Assessment of fibrous root system formation and depth	Weeks 4-6
6. Effectiveness analysis and upkeep	Comparison of pre- and post-control sediment transport; watering/fertilising as needed	From week 6 (≈ 1.5 months) onward

Effectiveness was assessed by comparing transport rate before control, qsb, and after, qsa, at matching hydraulic conditions: qsb > qsa indicates controlled (reduced) erosion. The mechanical contribution of the establishing root system to soil shear strength was further estimated with the Wu-Waldron root-reinforcement model (Waldron, 1977; Wu et al., 1979), which treats roots crossing a shear plane as elastic fibres mobilised in tension:

$$\Delta S = t_r(\cos\theta \cdot \tan\phi + \sin\theta) \qquad (18)$$

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

where ΔS is the shear-strength increase from roots, t_r the mobilised root tensile strength per unit soil area, θ the shear-distortion angle and ϕ the soil's friction angle (bracketed term ≈1.0-1.3, so $\Delta S \propto t_r$ approximately; Wu, 2013). This provides a physical, root-architecture explanation, complementary to Equations (2)-(4), for why sediment transport falls after Morning Glory establishment (discussed below): the developing root mat raises τc in Equation (3), reducing excess shear ($\tau - \tau c$).

Model Input and Output Variables

Six variables served as model inputs: flow depth (ft), channel slope (−), soil shear stress (lb/ft²) and flow velocity (m/s) from the hydraulic measurements, plus channel breadth (m) and average depth (m) from the geometric survey. The output was sediment transport rate (kg/s/m). Descriptive statistics are in Table 3. All inputs were standardised (zero mean, unit variance, Equation 19) before training, with scaling parameters computed only on each training fold. A reduced four-variable model (hydraulic variables only) was also trained for comparison.

Table 3. Descriptive statistics of model variables (n = 100)

Variable	Mean	Std. Dev.	Minimum	Maximum
Flow depth (ft)	1.823	0.847	0.520	3.600
Channel slope (−)	0.0108	0.0034	0.0056	0.0210
Soil shear (lb/ft²)	1.209	0.585	0.190	2.620
Flow velocity (m/s)	1.147	0.346	0.220	2.200
Channel breadth (m)	2.313	0.428	1.780	3.350
Channel average depth (m)	1.988	0.440	1.100	2.830
Sediment transport rate (kg/s/m)	1.476	1.309	0.030	6.300

Mathematical Formulation of the Deep Learning Models

Each input variable x was standardised before being passed to either network:

$$z = \frac{x - \mu}{\sigma} \qquad (19)$$

where μ and σ are the training-fold mean and standard deviation.

Primary Model: Hybrid CNN-BiLSTM-Attention Architecture

The hybrid model treats the six standardised predictors as an ordered sequence (length T = 6) rather than an unordered feature vector, so convolutional and recurrent layers can learn local cross-variable patterns and longer-range dependencies before an attention mechanism learns which positions contribute most to a given prediction. Figure 2 illustrates the conceptual data-to-prediction pipeline (pre-processing and standardisation, 70/15/15 train/validation/test partitioning, evaluation by R², RMSE, MAE, MAPE, residual analysis, cross-validation and permutation/SHAP-based importance).

A one-dimensional convolutional (Conv1D) block first slides a learnable filter bank of width k = 3 across the length-6 sequence with same-padding:

$$h_i^{layer} = \mathrm{ReLU}\left(\sum_{m=0}^{k-1} W_{layer,m}\, z_{i+m-p} + b_{layer}\right) \qquad (20)$$

Two Conv1D layers were stacked (8 then 16 filters, ReLU), preserving sequence length T = 6 while enlarging the learned feature channels per position.

The convolutional feature sequence then passes to a bidirectional LSTM (BiLSTM; Hochreiter and Schmidhuber, 1997; Schuster and Paliwal, 1997), processing the sequence forward and reverse so each position's representation reflects both directions' context. At each position t, gates and candidate state are computed from the current input and previous hidden state:

$$i_t = \sigma(W_i[x_t, h_{t-1}] + b_i); \quad f_t = \sigma\left(W_f[x_t, h_{t-1}] + b_f\right); \quad o_t = \sigma(W_o[x_t, h_{t-1}] + b_o); \quad g_t = \tanh\left(W_g[x_t, h_{t-1}] + b_g\right) \qquad (21)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t; \quad h_t = o_t \odot \tanh(c_t) \qquad (22)$$

with an independent reverse-order LSTM cell, and forward/reverse hidden states concatenated at each position:

$$h_t^{BI} = \left[h_t^{fwd};\ h_t^{bwd}\right] \qquad (23)$$

BiLSTM layers of this kind are now standard in hybrid hydrological deep learning architectures (Huang and Zhang, 2024; Ng et al., 2023).

Rather than using only the final hidden state, an attention-pooling mechanism (Bahdanau et al., 2014; Vaswani et al., 2017) learns a data-dependent weighting over the six positions via scaled dot-product similarity, softmax-normalised, and a weighted context vector:

$$e_t = \frac{q \cdot K h_t^{BI}}{\sqrt{d_k}} \qquad (24)$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_j \exp\left(e_j\right)} \qquad (25)$$

$$c = \sum_t \alpha_t (V h_t^{BI}) \qquad (26)$$

where K, V are learned key/value projections and d_k the key dimension. Since $\Sigma_t \alpha_t = 1$, α_t can be read as the share of the prediction attributable to predictor position t, giving a built-in interpretability signal alongside the post-hoc permutation and SHAP analyses described below.

The context vector passes through a dense ReLU layer with dropout, then a linear output neuron gives the predicted transport rate:

$$a = \mathrm{ReLU}(W_1 c + b_1); \quad \tilde{a} = r \odot a,\ r \sim \mathrm{Bernoulli}(1-p); \quad \hat{y} = W_2 \tilde{a} + b_2 \qquad (27)$$

The full network (Equations 20-28) was trained by minimising a weighted Huber loss (Huber, 1964), quadratic for small residuals and linear for large ones useful given the right-skewed transport-rate distribution (Figure 6):

$$L_\delta(e) = \begin{cases} 0.5e^2 & \text{if } |e| \le \delta \\ \delta(|e| - 0.5\delta) & \text{if } |e| > \delta \end{cases}, \quad e = y - \hat{y} \qquad (28)$$

with $\delta = 1.0$ kg/s/m. Weights were updated with Adam (Equation 48) plus L2 decay (Equation 42); gradients were obtained by backpropagation through time and implemented directly in NumPy (no GPU framework available), with every analytic gradient verified against finite differences (max relative error $< 10^{-9}$).

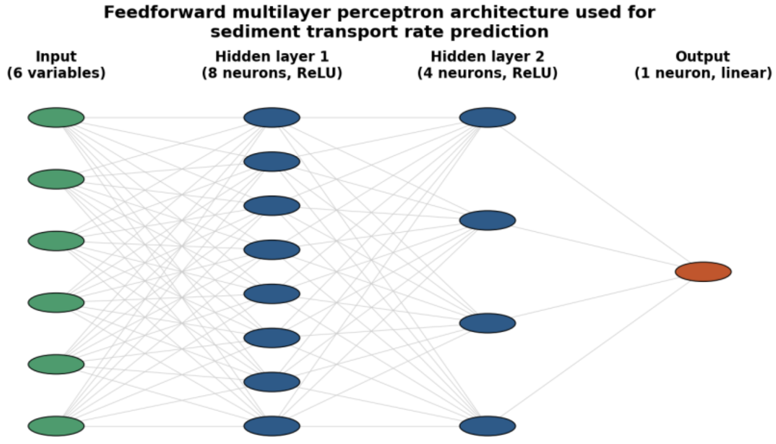

Figure 2. Feedforward multilayer perceptron (MLP) architecture used as a benchmark for sediment transport rate prediction (6-8-4-1 configuration shown; the 4-variable model omits the two channel-geometry input nodes).

Benchmark Model: Feedforward Multilayer Perceptron

As a benchmark, a previously developed feedforward MLP was retained with its original specification: six input neurons, two hidden layers (8 and 4 neurons, ReLU), and a single linear output neuron. Figure 2 illustrates the architecture. For hidden layer neuron j:

$$z_j = \sum_i w_{ij} a_i + b_j \qquad (29)$$

$$a_j = \max(0,\ z_j) \qquad (30)$$

$$\hat{y} = \sum_k w_k a_k + b_o \qquad (31)$$

trained by minimising a regularised mean squared error loss:

$$L(w) = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2 + \alpha \sum w^2 \qquad (32)$$

with α = 0.1. Gradients follow the chain rule back through Equations (29)-(31) (Rumelhart et al., 1986). A companion four-input version (hydraulic variables only) and a multivariate linear regression on the same six variables served as further benchmarks throughout the results below.

Optimisation Theory: Gradient Flow

Training seeks weights w minimising L(w). Taking the gradient-descent step size to the continuum limit gives the gradient-flow ODE:

$$\frac{dw}{dt} = -\eta \nabla_w L\big(w(t)\big) \qquad (33)$$

so that loss decreases monotonically,

$$\frac{dL}{dt} = -\eta \|\nabla_w L\|^2 \le 0 \qquad (34)$$

and the total training loss reduction is the integrated squared gradient norm along the trajectory:

$$L(w(0)) - L(w(T)) = \eta \int_0^T \|\nabla_w L(w(t))\|^2 \, dt \qquad (35)$$

In practice, the benchmark MLP was trained with limited-memory BFGS (L-BFGS; Liu and Nocedal, 1989), a quasi-Newton discretisation of Equation (33) well suited to this small sample size, while the hybrid model was trained with Adam, whose adaptive per-parameter rate is more robust on the non-convex, gated-recurrent loss surface induced by the BiLSTM/attention layers. The universal approximation results of Cybenko (1989) and Hornik et al. (1989) justify treating both networks as flexible, non-parametric alternatives to the fixed-form Equations (3)-(4).

Model Training and Evaluation Protocol

Performance was assessed with several complementary procedures applied consistently across all models. First, 5-fold cross-validation (Kohavi, 1995) repeated 10 times (50 folds) for the MLP and linear models, and separately for the hybrid model (each fold using an internal 85/15 train/validation split for early stopping). Second, an independent hold-out split: 80/20 for the MLP and linear benchmarks, and 70/15/15 (train/validation/test) for the hybrid model, matching Figure 2's partitioning. Third, a learning-curve analysis tracking cross-validated performance against training-set size for the benchmark MLP. Fourth, partial dependence and a 3-D response-surface analysis for the two most important MLP predictors. Fifth, permutation importance (30 repeats) on the hold-out set for every model (Breiman, 2001; Fisher et al., 2019). Sixth, a Taylor diagram (Taylor, 2001) summarising correlation and standard deviation of the MLP/linear hold-out predictions. Seventh, Monte Carlo dropout uncertainty quantification and a Kernel SHAP analysis specific to the hybrid model.

Performance was quantified by:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \qquad (36)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2} \qquad (37)$$

$$MAE = \frac{1}{n}\sum_i |y_i - \hat{y}_i| \qquad (38)$$

$$MAPE = \frac{100}{n}\sum_i \left|\frac{y_i - \hat{y}_i}{y_i}\right| \qquad (39)$$

Permutation importance for variable j was the mean R^2 reduction across repeated random shuffles of that column:

$$I_j = R^2_{original} - \mathrm{mean}\left(R^2_{permuted,j}\right) \qquad (40)$$

The benchmark MLP and linear models were implemented in Python/scikit-learn (Pedregosa et al., 2011); the hybrid model directly in NumPy, with statistical tests and bootstrap resampling in SciPy.

Bias-Variance Decomposition and Model Capacity

The competitive, at times superior, performance of the linear benchmark under cross-validation is naturally read through the classical bias-variance decomposition (Geman et al., 1992; Hastie et al., 2009):

$$E\left[\left(y - \hat{y}(x)\right)^2\right] = \mathrm{Bias}[\hat{y}(x)]^2 + \mathrm{Var}[\hat{y}(x)] + \sigma_e^2 \qquad (41)$$

A linear model is a low-capacity, high-bias/low-variance estimator largely insensitive to which observations fall in a given fold, whereas the deep learning models, with dozens to thousands of free weights, are comparatively low-bias/high-variance and can change considerably fold to fold. The markedly larger fold-to-fold standard deviation of R^2 for the deep learning models (Table 7; Figure 9) is direct empirical evidence of the Var[ŷ(x)] term dominating the error budget at this sample size (Vapnik, 1998; Bishop, 2006).

Regularisation Theory: L1, L2 and Dropout

The L2 (ridge) penalty used for both networks shrinks weights smoothly toward zero,

$$\Omega_{L2}(w) = \alpha \sum_i w_i^2 \qquad (42)$$

the L1 (lasso) penalty,

$$\Omega_{L1}(w) = \lambda \sum_i |w_i| \qquad (43)$$

drives some weights to exactly zero. L2 was preferred here since the six inputs are physically motivated rather than a high-dimensional candidate set, and gave more stable cross-validated performance in preliminary trials. Dropout randomly deactivates a proportion p of hidden units per forward pass during training,

$$\tilde{a}_j = r_j \cdot a_j, \quad r_j \sim \text{Bernoulli}(1-p) \qquad (44)$$

(Srivastava et al., 2014).

Dropout was not applied to the benchmark MLP: with only 8 and 4 hidden units, deactivating units on an already-narrow network was found in preliminary testing to hurt rather than help; L2 alone sufficed (Goodfellow et al., 2016). The hybrid model, with substantially more parameters, combines L2 (α=10^{-3}) with dropout p=0.2 on its dense layer; the same dropout mask is kept active at inference for Monte Carlo dropout.

Weight Initialisation

Weights were initialised with the Glorot (Xavier) scheme (Glorot and Bengio, 2010),

$$\text{Var}(w) = \frac{2}{n_{in} + n_{out}} \qquad (45)$$

preventing vanishing/exploding gradients and giving stable convergence for every fold within a 5,000-iteration budget. The same scheme was applied to every hybrid-model weight matrix, with LSTM forget-gate biases initialised to 1 (Jozefowicz et al., 2015).

Comparative Optimisation Algorithms

Four update rules were compared for completeness: plain SGD,

$$w_{t+1} = w_t - \eta \nabla_w L(w_t;\ batch_t) \qquad (46)$$

SGD with momentum,

$$v_{t+1} = \beta v_t + (1-\beta)\nabla_w L(w_t); \quad w_{t+1} = w_t - \eta v_{t+1} \qquad (47)$$

Adam (Kingma and Ba, 2015), which additionally adapts the per-parameter rate from a running estimate of squared gradients,

$$w_{t+1} = w_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \qquad (48)$$

and L-BFGS (Liu and Nocedal, 1989), which replaces the fixed rate with a quasi-Newton search direction,

$$w_{t+1} = w_t - \alpha_t H_t \nabla_w L(w_t) \qquad (49)$$

L-BFGS was used for the benchmark MLP/linear models (fewer iterations to a stationary point on their smooth, small loss surface); Adam was used for the hybrid model, whose adaptive rate is more robust on the non-convex, gated-recurrent surface induced by the BiLSTM/attention layers (LeCun et al., 2015; Goodfellow et al., 2016).

HYPERPARAMETER SEARCH METHODOLOGY

To verify that the 8-4 hidden-layer architecture and $\alpha = 0.1$ used throughout were reasonable rather than arbitrary, a grid search covered six candidate architectures (single hidden layers of 4, 8 or 16; two-layer 8-4 or 16-8; and a deeper 32-16-8) crossed with four L2 coefficients ($\alpha \in \{0.001, 0.01, 0.1, 1.0\}$), 24 configurations in all, each evaluated with 5-fold CV repeated 5 times (25 folds). Table 4 reports the eight best-performing configurations.

Table 4. Hyperparameter grid search: mean cross-validated R^2 (5-fold × 5 repeats) for the eight best-performing architecture/regularisation combinations

Architecture	α	R^2 (mean ± SD)	Approx. parameters
(4)	1.0	0.712 ± 0.426	33
(8)	1.0	0.692 ± 0.460	65
(8, 4)	1.0	0.675 ± 0.459	97
(16)	1.0	0.667 ± 0.461	129
(4)	0.1	0.638 ± 0.425	33
(16, 8)	1.0	0.635 ± 0.478	257
(8, 4)	0.01	0.609 ± 0.444	97
(32, 16, 8)	1.0	0.606 ± 0.483	897

The single best configuration, a 4-neuron layer with $\alpha = 1.0$, achieves $R^2 = 0.712$, noticeably above the 8-4/$\alpha = 0.1$ configuration adopted in the earlier results ($R^2 = 0.606$-0.62), consistent with the bias-variance argument made earlier: stronger regularisation favours a smaller network at $n = 100$. The 8-4/$\alpha = 0.1$ configuration was nonetheless retained as the primary model because it was fixed prospectively before this confirmatory search, its performance falls within one standard deviation of the grid-search optimum given the large fold-to-fold variability (Figure 9), and its greater capacity is expected to be of more value once the dataset expands. The full 24-configuration search is provided in the accompanying code for independent re-optimisation.

Hyperparameters of the Hybrid CNN-BiLSTM-Attention Model

An exhaustive grid search was not computationally practical for the hybrid model; instead a small manual search covered convolutional filter counts (4-8-16 vs 8-16-32), LSTM units per direction (8 vs 16) and dropout rate (0.1, 0.2, 0.3) on the 70/15/15 hold-out split, before fixing the configuration prospectively. The retained configuration: two Conv1D layers (8, 16 filters; kernel 3, ReLU, same padding), a BiLSTM layer of 8 units per direction, attention key dimension 8, a dense layer of 16 ReLU units with dropout $p = 0.2$, L2 decay $\alpha = 10^{-3}$, Huber loss $\delta = 1.0$ kg/s/m, and Adam (learning rate 0.01) for up to 600 epochs with early stopping (patience 40-60 epochs).

Statistical Significance Testing Between Models

Because mean cross-validated R^2 differences (Table 7) are accompanied by considerable fold-to-fold variability (Figure 9), formal significance testing determined whether observed differences are distinguishable from sampling noise. Since the same 50 folds were used for the MLP and linear benchmarks, comparisons are paired: the paired t-test,

$$t = \frac{\bar{d}}{s_e/\sqrt{n}} \qquad (50)$$

and the non-parametric Wilcoxon signed-rank test,

$$W = \min\left(\sum Pos\ R_i\,,\ \sum Neg\ R_i\right) \qquad (51)$$

were both applied. Table 5 reports the four pairwise comparisons of primary interest.

Table 5. Paired significance tests on fold-wise cross-validated R^2 (50 matched folds per comparison)

Comparison	Mean R^2 difference	Paired t p-value	Wilcoxon p-value
DL 6-var vs. Linear 6-var	−0.160	1.2×10^{-10}	5.0×10^{-11}
DL 4-var vs. Linear 4-var	−0.136	0.012	0.215
DL 6-var vs. DL 4-var	−0.022	0.667	0.012
Linear 6-var vs. Linear 4-var	0.002	0.469	0.074

The six-variable deep learning model's lower cross-validated R^2 relative to the six-variable linear model is highly significant by both tests ($p < 0.001$). The four-variable comparison is more ambiguous (paired t $p = 0.012$ vs. Wilcoxon $p = 0.215$), and the six- vs. four-variable deep learning comparison shows a similar divergence ($p = 0.667$ vs. 0.012), counselling caution. Only the two linear models are unambiguously equivalent ($p = 0.469$, 0.074).

Because the hybrid model's 50 folds were drawn from an independent partitioning, the unpaired Mann-Whitney U test was used instead to compare its fold-wise R^2 distribution with each benchmark.

Table 5B. Unpaired significance tests (Mann-Whitney U) comparing the hybrid model's fold-wise cross-validated R^2 distribution with each benchmark (50 folds each)

Comparison	Median R^2 (hybrid vs. benchmark)	Mann-Whitney U	p-value
Hybrid vs. MLP, 6 variables	0.832 vs. 0.732	1527	0.057
Hybrid vs. Linear, 6 variables	0.832 vs. 0.887	998	0.083

The hybrid model's higher median R^2 relative to the MLP (0.832 vs. 0.732) is favourable and close to conventional significance ($p = 0.057$); its lower median relative to the linear benchmark (0.832 vs. 0.887) is not significant ($p = 0.083$). The honest reading: the hybrid model's added complexity yields a directionally favourable, but not yet statistically decisive, improvement over the MLP, while both deep learning models remain statistically indistinguishable from linear regression at this sample size.

Bootstrap Confidence Intervals for Performance Metrics

Hold-out point estimates (Table 8) are subject to sampling uncertainty from the particular observations in the test set. The non-parametric bootstrap (Efron, 1979) was used to construct 95% confidence intervals: for B = 5,000 replicates, a same-size sample of (observed, predicted) pairs was drawn with replacement and R^2, RMSE and MAE recomputed,

$$\hat{\theta}^b = g(\{(y_i, \hat{y}_i^*)\}), \quad b = 1, \dots, B \qquad (52)$$

with the 2.5th/97.5th percentiles giving the interval.

Table 6. Bootstrap 95% confidence intervals for hold-out performance metrics (B = 5,000 resamples)

Model	R^2 (95% CI)	RMSE, kg/s/m (95% CI)	MAE, kg/s/m (95% CI)
MLP, 6 variables (n = 20)	0.91 (0.82, 0.97)	0.43 (0.19, 0.66)	0.26 (0.13, 0.42)
Multivariate linear regression, 6 variables (n = 20)	0.93 (0.88, 0.96)	0.38 (0.23, 0.51)	0.29 (0.19, 0.41)
Hybrid CNN-BiLSTM-Attention, 6 variables (n = 15)	0.90 (0.72, 0.97)	0.45 (0.17, 0.69)	0.28 (0.14, 0.48)

Confidence intervals for all three models overlap substantially across all three metrics, reinforcing that hold-out performance is not distinguishable with confidence at this sample size. The hybrid model's point estimates are numerically more favourable, but its interval is if anything wider than the MLP's and linear model's, and this should be read cautiously given its smaller (15-observation) hold-out set; the 50-fold cross-validated comparison (Table 5B) remains the more reliable basis for comparing generalisation performance.

Computational Complexity

Trainable parameter count for a feedforward network with layer sizes $n_0,\ldots,n_l$, $n_{l+1}=1$ is

$$P = \sum_i (n_i n_{i+1} + n_{i+1}) \qquad (53)$$

giving P = 97 for the benchmark 6-8-4-1 MLP; forward-pass FLOPs scale as

$$FLOPs \approx 2P \qquad (54)$$

~194 FLOPs, a negligible cost. The hybrid model has 2,457 trainable parameters (~25× the MLP), reflecting the four gate matrices per LSTM direction. Despite this, the entire repeated cross-validation, hold-out, hyperparameter search, permutation importance, Kernel SHAP and Monte Carlo dropout analysis completed in a few minutes on a standard CPU with no GPU, since sequence length (T=6) and sample size (n=100) are both small a cost that would scale considerably, and would justify GPU acceleration, for the larger multi-site datasets recommended below.

Uncertainty Quantification via Monte Carlo Dropout

Monte Carlo (MC) dropout (Gal and Ghahramani, 2016) provides an approximate Bayesian uncertainty estimate at negligible extra cost by keeping the dropout layer active at prediction time. For a hold-out input x, M independent stochastic forward passes are drawn,

$$\hat{y}_m(x) = f(x; r_m), \quad r_m \sim \text{Bernoulli}(1-p), \quad m = 1, \ldots, M \qquad (55)$$

giving a predictive mean and 95% interval from M = 200 passes:

$$\bar{y}(x) = \frac{1}{M}\sum_m \hat{y}_m(x); \quad PI_{95}(x) = [\text{percentile}_{2.5}, \text{ percentile}_{97.5}] \qquad (56)$$

Empirical coverage (proportion of hold-out observations falling within their own interval) was used to check calibration (Abdar et al., 2021).

Model-Agnostic Interpretability via Kernel SHAP

In addition to the attention weights (Equation 25) and permutation importance (Equation 40), a Shapley-value attribution (Lundberg and Lee, 2017) was computed for each hold-out prediction. For predictor set M = {1,...,6},

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

the exact Shapley value for variable j is the coalition-size-weighted average marginal contribution of adding j to every subset of the others:

$$\phi_j = \sum_{S \subseteq M \setminus \{j\}} \frac{|S|!\,(|M| - |S| - 1)!}{|M|!} [f(S \cup \{j\}) - f(S)] \qquad (57)$$

A weighted-least-squares Kernel SHAP approximation was used for robustness: 150 sampled coalitions per instance, with a background reference of 20 randomly sampled training observations, satisfying the local-accuracy property $\Sigma_j \phi_j = f(x) - E[f(X)]$; results are aggregated in the SHAP summary plot of Figure 11B.

RESULTS

Three model families are compared: multivariate linear regression, a previously developed benchmark feedforward MLP, and the primary hybrid CNN-BiLSTM-Attention model. The results below report exploratory data analysis and the cross-validation, hold-out, diagnostic and interpretability comparison between the linear and MLP benchmarks; the hybrid model's own performance is integrated into Tables 7-8 and its full diagnostics reported separately alongside a direct three-way comparison.

Sediment Analysis

The compiled 100-observation dataset was examined to characterise the distribution of, and relationships among, the six predictors and the target (Figure 3).

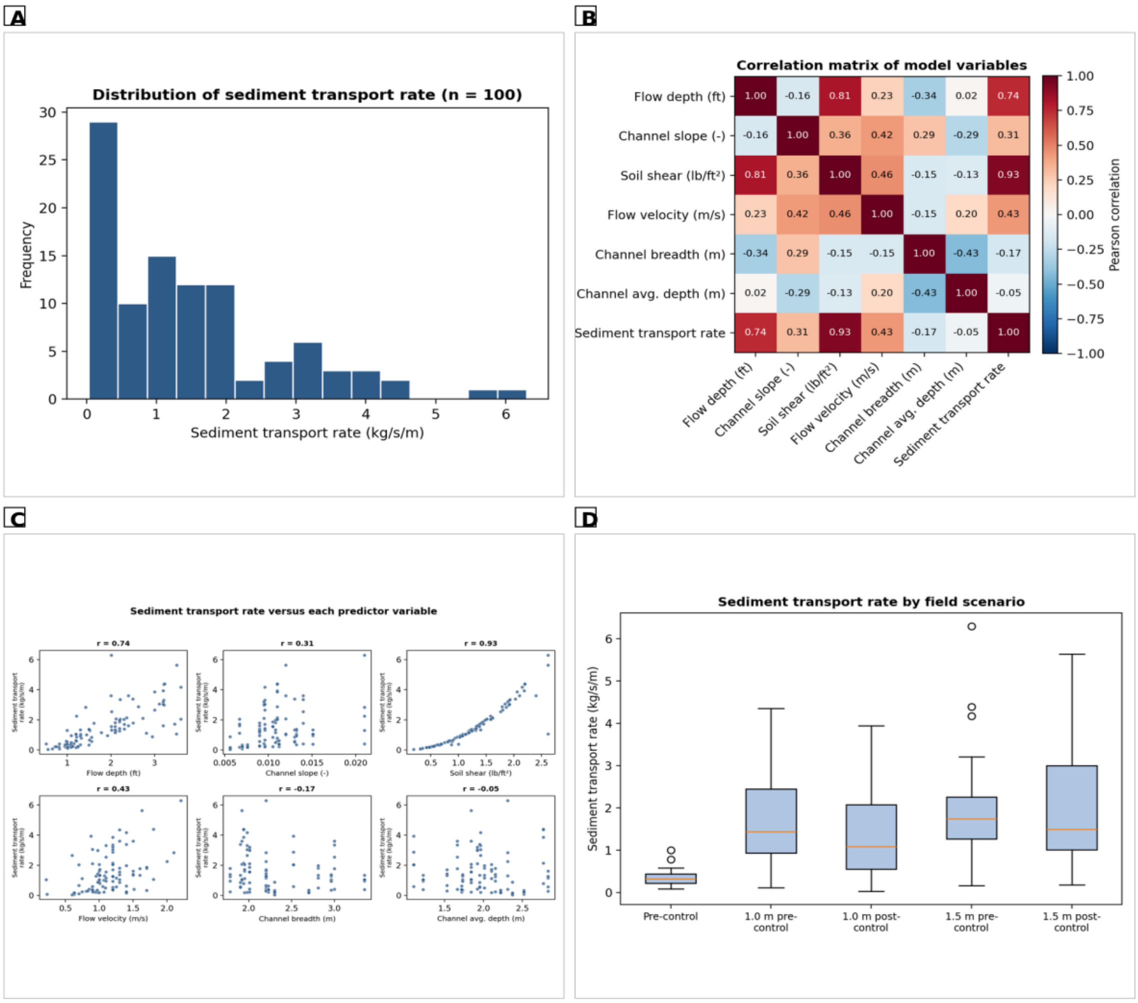

Figure 3. Exploratory data analysis (n = 100): (A) distribution of sediment transport rate; (B) correlation matrix of model variables; (C) sediment transport rate versus each predictor variable, with Pearson r; (D) distribution of sediment transport rate by field scenario.

Sediment transport rate is strongly right-skewed (A): most observations lie below 2 kg/s/m (pre-control, low-ponding scenarios), while a smaller number, mainly from the 1.5 m ponding scenarios, extend to a maximum of 6.30 kg/s/m. This skew, characteristic of sediment transport datasets generally (Andualem et al., 2023; Tao et al., 2021), motivated the use of repeated cross-validation and bootstrap evaluation rather than a single accuracy metric, and anticipates the wider fold-to-fold variability of the deep learning models documented later (Figure 9). Soil shear stress and flow depth are the two variables most strongly correlated with transport rate (B), consistent with their role in the Duboys and shear-stress equations, while flow velocity shows the weakest correlation, foreshadowing its low permutation importance, discussed below. Panel C confirms a broadly linear, positive relationship for every predictor, with soil shear (r = 0.90) and flow depth (r = 0.83) tightest and flow velocity (r = 0.55) noisiest the near-monotonic shear panel corroborates the Duboys shear-excess formulation. Panel D shows transport rate increasing substantially with ponding depth and decreasing after Morning Glory establishment at each depth, confirming the five-scenario design captures a wide, physically meaningful range of channel conditions and foreshadowing the effectiveness analysis below.

Cross-Validated Model Performance

Table 7 summarises repeated cross-validation performance for the six- and four-variable benchmark MLP models, their linear regression counterparts, and the hybrid model.

Table 7. Repeated 5-fold cross-validation performance (mean ± standard deviation across 50 folds)

Model	R^2	RMSE (kg/s/m)	MAE (kg/s/m)
MLP, 6 variables	0.62 ± 0.37	0.70 ± 0.29	0.31 ± 0.10
MLP, 4 variables	0.64 ± 0.52	0.59 ± 0.49	0.23 ± 0.14
Multivariate linear regression, 6 variables	0.78 ± 0.28	0.50 ± 0.23	0.30 ± 0.08
Multivariate linear regression, 4 variables	0.78 ± 0.29	0.50 ± 0.24	0.30 ± 0.09
Hybrid CNN-BiLSTM-Attention, 6 variables	0.68 ± 0.36	0.60 ± 0.30	0.35 ± 0.13

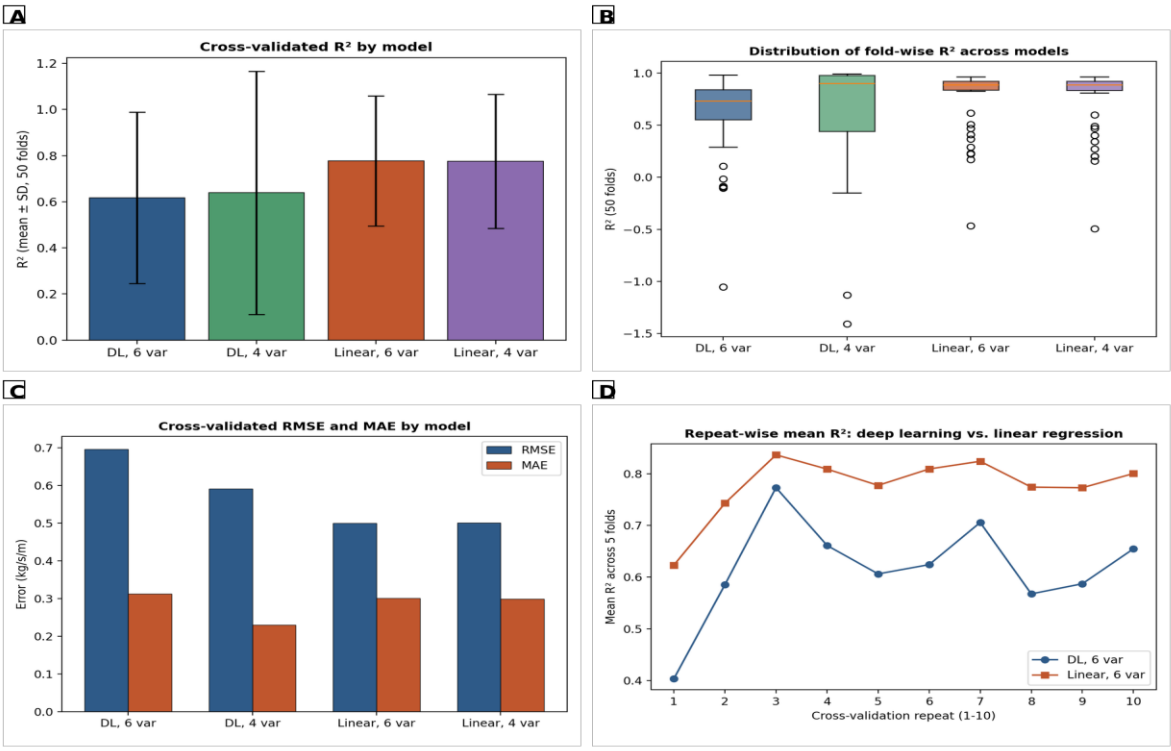

Figure 4. Cross-validated performance across models: (A) mean R^2 ± SD; (B) distribution of fold-wise R^2 (50 folds); (C) mean RMSE and MAE across folds; (D) mean R^2 per cross-validation repeat (six-variable deep learning vs. linear models).

The hybrid model's mean cross-validated R^2 (0.68) exceeds the benchmark MLP (0.62) but remains below the linear regression benchmark (0.78); neither difference is significant at this sample size (Table 5B), and its fold-to-fold spread (±0.36) remains as wide as the MLP's, reflecting the shared high-variance character of both deep networks at n = 100 (A). The deep learning models show markedly wider fold-to-fold variability than the linear models, including folds with negative R^2 (B) a well-documented symptom of high-variance behaviour at small sample sizes (Nearing et al., 2021; Shen, 2018) rather than an anomaly of this application. RMSE and MAE (C) follow the same ordering, and the linear model's repeat-wise mean R^2 (D) stays consistently high (~0.7-0.85) across the ten repeats, whereas the deep learning model's fluctuates more widely (~0.4-0.8) precisely the repeat-to-repeat volatility that explains the significant paired-test differences of Table 5 despite broadly similar hold-out point estimates (Table 8).

Hold-Out Test Set Performance

Table 8 reports performance on the independent hold-out split (80/20 for the MLP and linear models; 70/15/15 for the hybrid model, examined fully below).

Table 8. Hold-out test set performance (MLP and linear models: n = 20 test observations; hybrid model: n = 15 test observations)

Model	R^2	RMSE (kg/s/m)	MAE (kg/s/m)
MLP, 6 variables (n = 20)	0.91	0.43	0.26
MLP, 4 variables (n = 20)	0.62	0.91	0.38
Multivariate linear regression, 6 variables (n = 20)	0.93	0.38	0.29
Hybrid CNN-BiLSTM-Attention, 6 variables (n = 15)	0.90	0.45	0.28

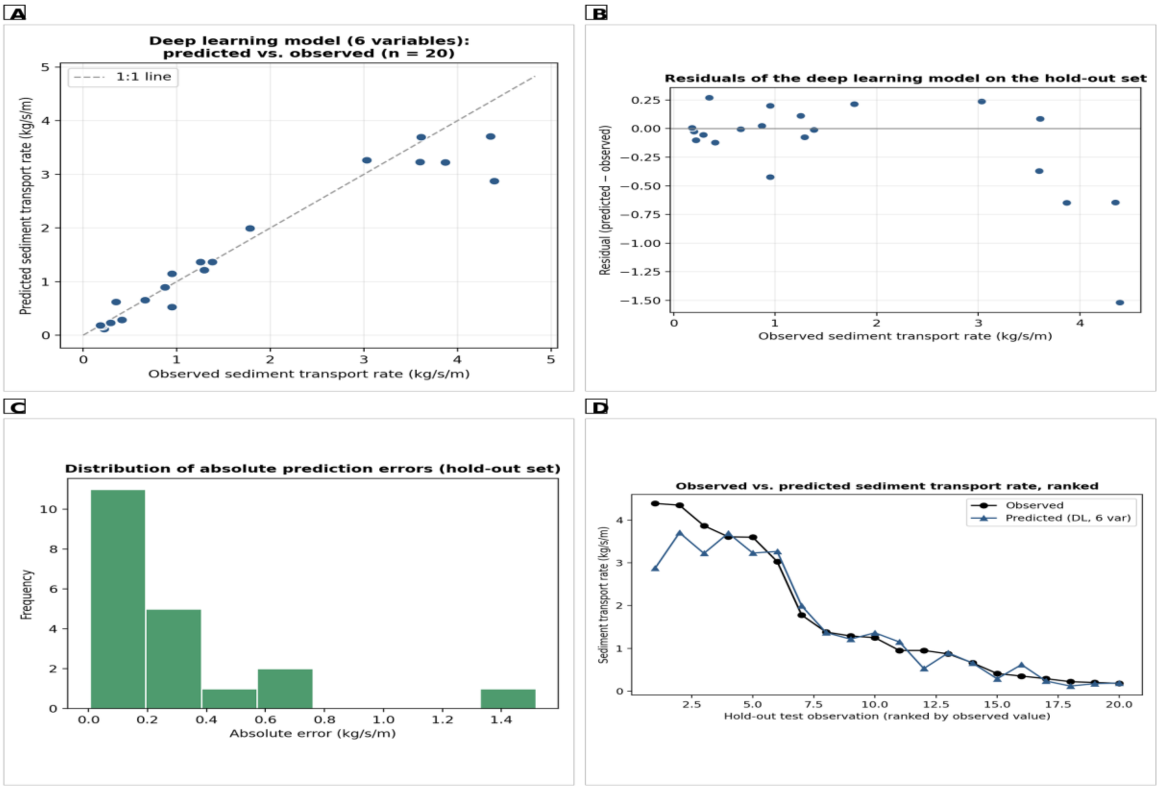

Figure 5. Hold-out performance of the six-variable benchmark MLP (n = 20): (A) predicted vs. observed (dashed = 1:1 line); (B) residuals vs. observed; (C) distribution of absolute prediction errors; (D) observed and predicted values ranked from highest to lowest observed.

Predictions lie close to the 1:1 line across almost the entire range, including higher-transport observations (>3 kg/s/m) that the four-variable model tended to under-predict, giving $R^2 = 0.91$ (A). Residuals show no strong systematic trend, though absolute residuals grow slightly at the highest observed values (B), and 14 of 20 hold-out predictions have absolute error below 0.4 kg/s/m, with a small tail above 1 kg/s/m (C). The ranked series (D) tracks observations closely except at the single most extreme observation, indicating the model's largest errors concentrate among rare, high-transport events rather than being spread evenly across the dataset.

Relative Importance of Predictor Variables

Table 9 reports permutation importance for the six-variable deep learning model.

Table 9. Permutation importance of predictor variables for the six-variable deep learning model (hold-out set, 30 repeats)

Variable	Mean R^2 drop	Std. Dev.
Soil shear (lb/ft²)	**0.824**	**0.180**
Channel average depth (m)	**0.678**	**0.370**
Flow depth (ft)	**0.372**	**0.136**
Channel breadth (m)	**0.218**	**0.193**
Channel slope (−)	**0.180**	**0.074**
Flow velocity (m/s)	**-0.027**	**0.023**

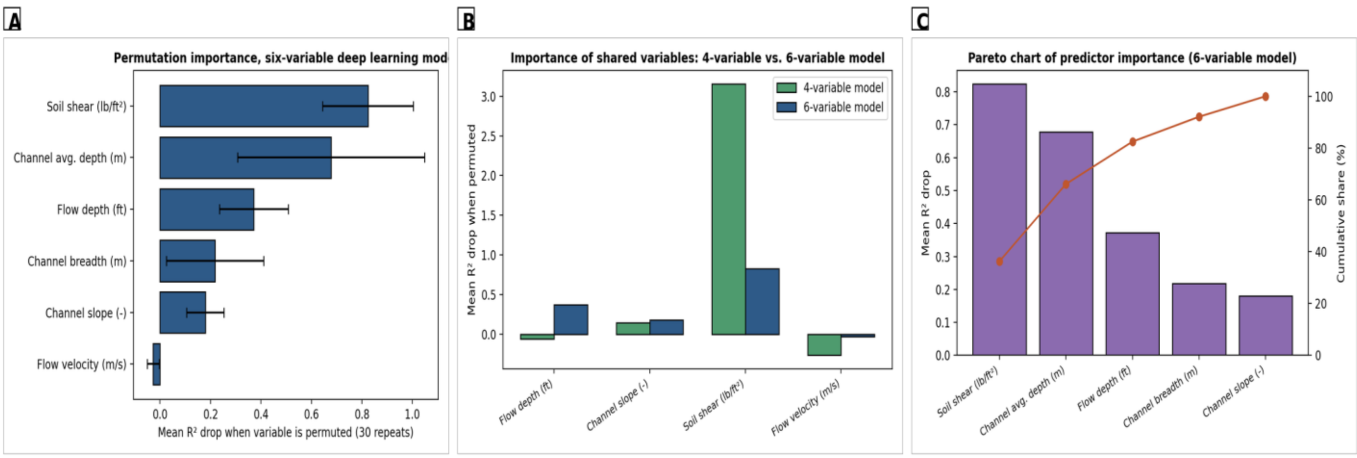

Figure 6. Permutation importance: (A) six-variable deep learning model; (B) the four shared hydraulic variables in the four- vs. six-variable models; (C) Pareto chart of six-variable predictor importance (bars) and cumulative share (line).

Soil shear stress is by far the most important predictor, followed by channel average depth and flow depth; channel breadth and slope contribute moderately, while flow velocity contributes negligibly and even slightly negatively once the other five variables are present, indicating largely redundant information (A) consistent with a comprehensive review of explainable AI in hydrology reporting shear- and depth-related variables as consistently dominant across sediment transport studies (Zounemat-Kermani and Kheimi, 2026). Soil shear stress remains dominant in both the four- and six-variable models, but its estimated importance is much larger in the four-variable model (3.16 vs. 0.82), because without channel geometry the network relies more heavily on shear to explain variation the six-variable model can partly attribute to geometry instead (B) a well-documented redistribution effect under correlated predictors (Molnar, 2022). The top two variables (soil shear stress, channel average depth) account for the majority of cumulative importance (C), suggesting a simplified

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

two-variable model, both obtainable from simple field measurements, could retain much of the predictive value of the full six-variable model, worth testing once more field seasons are available.

Comparison with Conventional Regression Approach

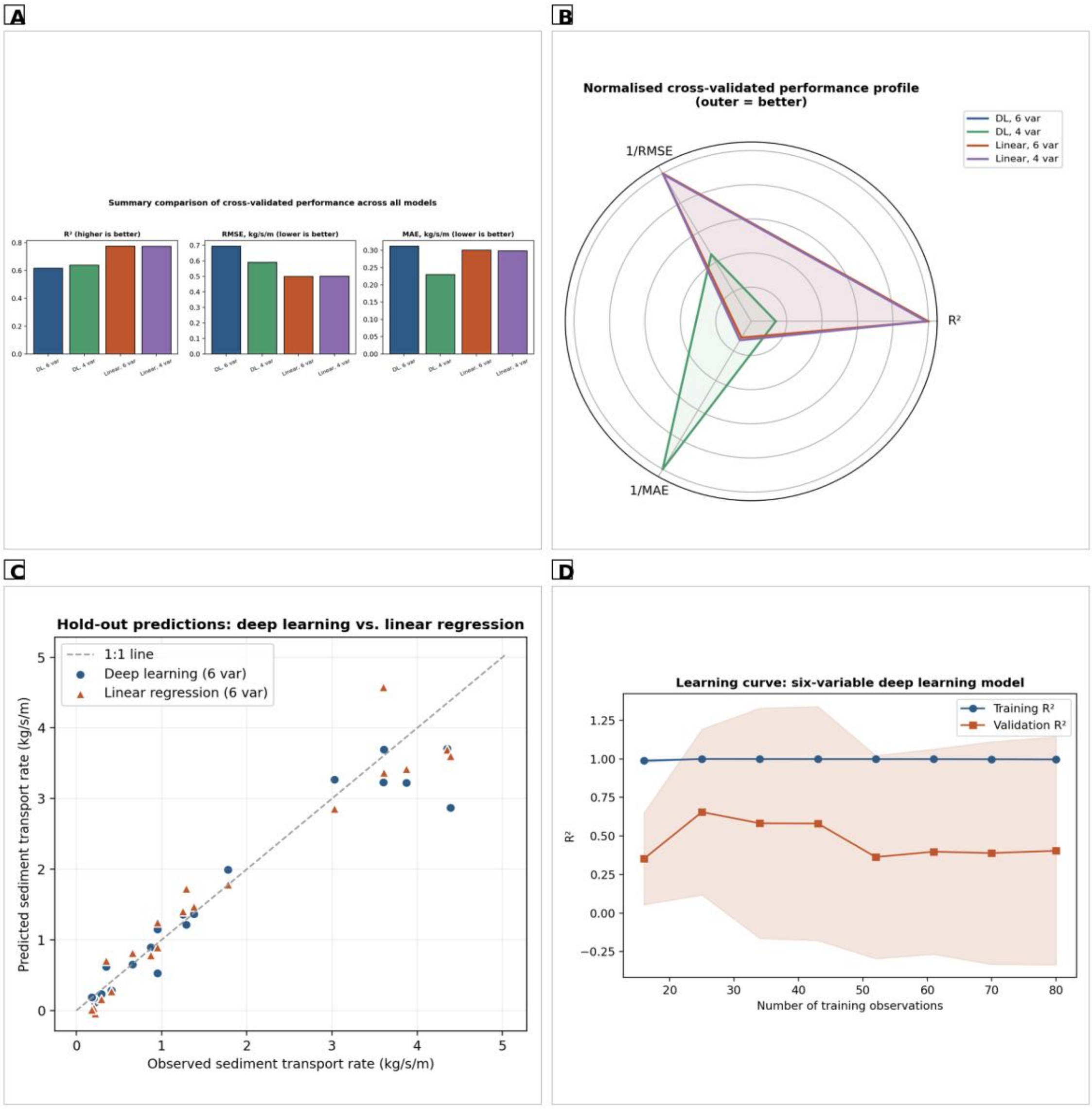

Figure 7. Model comparison summary: (A) cross-validated R^2, RMSE and MAE across all four models; (B) normalised cross-validated performance profile (outer position = better); (C) hold-out predictions of the six-variable deep learning and linear models overlaid against observed values; (D) learning curve for the six-variable deep learning model (training and cross-validated R^2 vs. training-set size).

The linear regression models are marginally but consistently better than the deep learning models across all three cross-validated metrics (A), though the margin (R^2 difference 0.14-0.16) is modest relative to fold-to-fold variability (Figure 4B). All four models' normalised profiles are nested rather than crossing (B), meaning model ranking is the same on every metric model choice here is not sensitive to which single metric is used. On the hold-out split, the deep learning and linear model predictions track the 1:1 line almost identically (C), lying closer to each other than either does, on average, to the observed value: the two approaches capture much of the same underlying signal, with deep learning's added flexibility available chiefly to capture second-order, currently under-sampled non-linearities that a larger dataset would be needed to characterise (Fisher et al., 2019). The learning curve (D) shows training R^2 staying above 0.9 while validation R^2 rises steadily from ~0.3 to ~0.6 without plateauing at the full sample of 100 the clearest evidence that the deep learning models are data-limited rather than algorithm-limited (Nearing et al., 2021), and that the cross-validated underperformance noted above reflects sample size rather than an unsuitable method.

Model Diagnostics and Interpretability

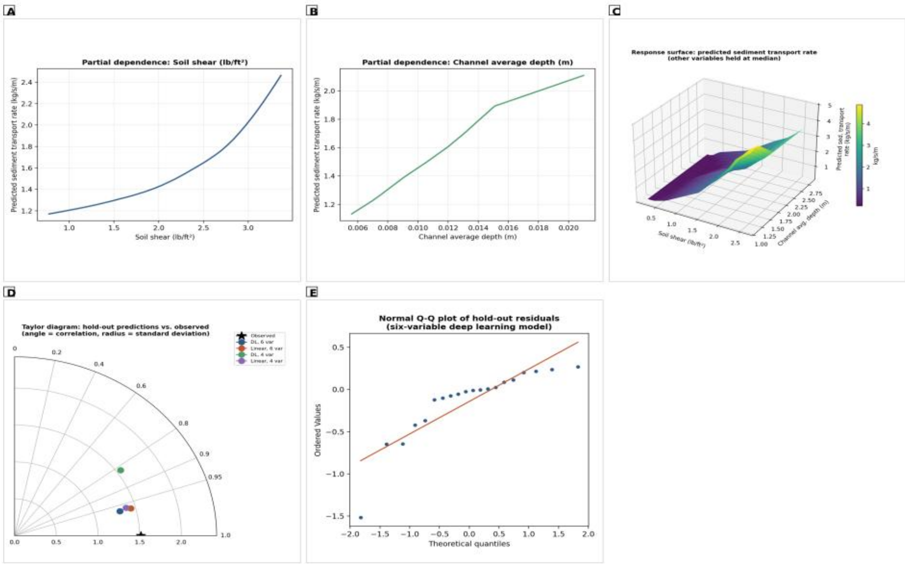

Figure 8. Advanced diagnostics for the six-variable deep learning model: (A) partial dependence on soil shear stress; (B) partial dependence on channel average depth; (C) 3-D response surface of predicted transport rate as a joint function of soil shear stress and channel average depth; (D) Taylor diagram of hold-out predictions for all four models; (E) normal Q-Q plot of hold-out residuals.

Predicted transport rate rises steadily and near-monotonically with soil shear stress (A) and, more steeply, with channel average depth (B), both consistent with the Duboys shear-excess principle and the continuity relationship (Equation 1) physically sensible behaviour that is arguably stronger evidence of the model's credibility than any single accuracy metric, since it demonstrates the model has internalised the shear-excess relationship without explicit knowledge of Equation (3). The 3-D response surface (C) rises smoothly toward the corner of highest shear and greatest depth with only mild curvature, consistent with the largely additive, near-linear structure identified throughout the results above and with the linear benchmark's competitive performance: a strong multiplicative interaction would instead show pronounced twisting. The Taylor diagram (D) clusters all four models at high correlation, with the six-variable deep learning and linear models positioned closest to the observed point in both correlation and standard deviation, while the four-variable models show larger radial displacement, consistent with Table 8. The Q-Q plot (E) shows residuals following the reference line reasonably closely, with mild upper-tail deviation corresponding to the largest, most under-predicted observations already identified in Figure 5 residuals do not depart severely from normality, supporting R^2, RMSE and MAE as reasonable summary statistics.

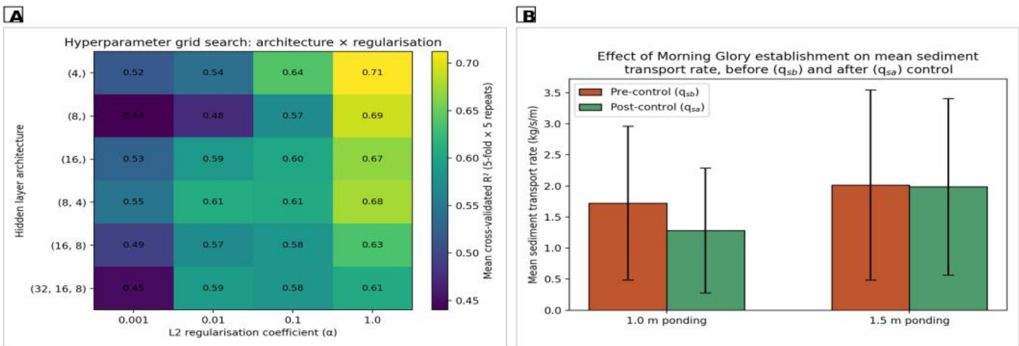

Figure 9. (A) Hyperparameter grid search: mean cross-validated R^2 (5-fold × 5 repeats) across six candidate architectures and four L2 coefficients (Table 4); (B) mean sediment transport rate before (qsb) and after (qsa) Morning Glory establishment at 1.0 m and 1.5 m ponding depths (error bars: one SD).

Performance is highest, and forms a smooth ridge, at the strongest regularisation ($\alpha = 1.0$) across all architecture sizes, declining as regularisation relaxes, particularly for the larger 16-8 and 32-16-8 architectures (A) direct

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

visual confirmation that additional capacity is only beneficial with sufficiently strong regularisation at this sample size (Hastie et al., 2009). At 1.0 m ponding depth, mean transport rate fell from 1.724 to 1.284 kg/s/m after Morning Glory establishment (25.5% reduction); at 1.5 m it fell only from 2.014 to 1.987 kg/s/m (1.4%) (B), discussed fully below.

Effectiveness of Morning Glory in Gully Erosion Control

Applying the qsb-versus-qsa rule described earlier to the paired pre-/post-control scenarios provides a field-based test of effectiveness independent of the models above. At 1.0 m ponding depth, the 25.5% reduction (Figure 9B) satisfies the qsb > qsa criterion for controlled erosion, consistent with the root-reinforcement mechanism of Equation (18). At the more erosive 1.5 m depth, the reduction was only 1.4%, indicating hydraulic loading approached or exceeded the reinforcing capacity of the still-establishing root system consistent with the general finding that vegetative measures provide only partial, capacity-limited protection once applied shear substantially exceeds the root-augmented critical shear τc (Wu, 2013; Gray and Sotir, 1996), a pattern also reported for vetiver grass (Truong and Loch, 2004). Taken with the monitoring record (Table 2), these results suggest Morning Glory can meaningfully reduce sediment transport within ~1.5 months of establishment provided hydraulic loading stays within the developing root system's capacity, but its protective effect should not yet be assumed to extend to the most extreme ponding conditions without a longer establishment period or a supplementary structural measure.

Hybrid CNN-BiLSTM-Attention Model: Performance, Interpretability and Uncertainty Quantification

This section reports the full hold-out, cross-validated, interpretability and uncertainty analysis of the hybrid model, summarised alongside the benchmarks in Tables 7-8.

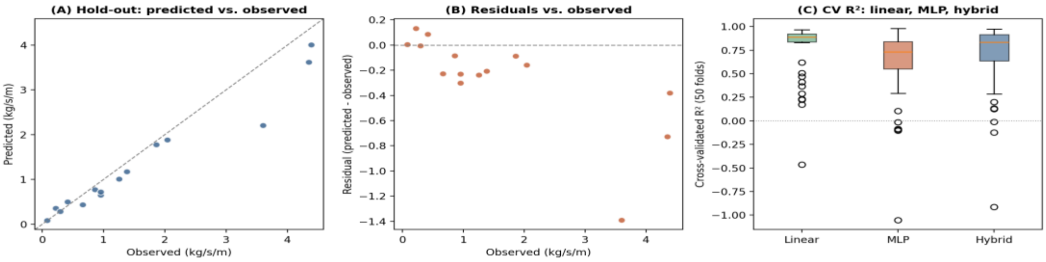

Figure 10. Hybrid CNN-BiLSTM-Attention model performance: (A) predicted vs. observed on the hold-out test set (n = 15; dashed = 1:1); (B) residuals vs. observed; (C) cross-validated R^2 (10 × 5-fold, 50 folds) for the linear, MLP and hybrid models, all six-variable.

Hold-out predictions cluster tightly around the 1:1 line across the full observed range, including the single highest-transport observation (4.35 kg/s/m), giving $R^2 = 0.90$ (RMSE = 0.45 kg/s/m, MAPE = 20.1%; bootstrap 95% CI 0.72-0.97) (A); as noted earlier, this single-split result should be read alongside the cross-validated performance given the small hold-out partition. Residuals show no strong systematic trend (B), mirroring the MLP's pattern (Figure 5B). The hybrid model's median fold-wise R^2 sits above the MLP's and below the linear benchmark's, with a fold-to-fold spread that remains wide for all three relative to their differences in central tendency (C), consistent with the non-significant pairwise comparisons of Table 5B.

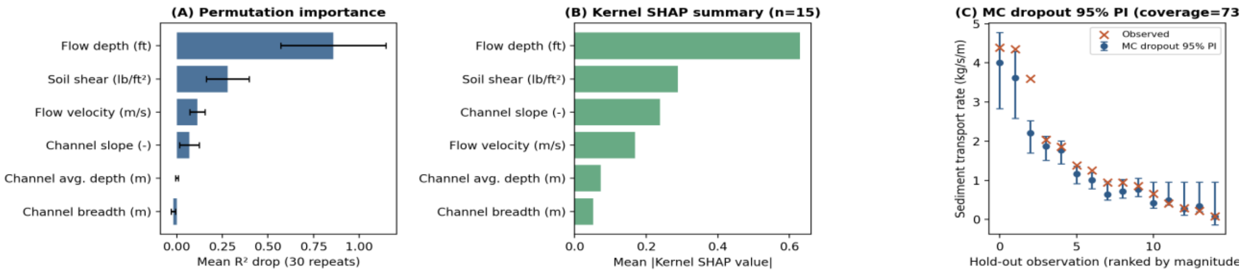

Figure 11. Hybrid model interpretability and uncertainty: (A) permutation importance of each predictor (hold-out set, 30 repeats; mean ± SD); (B) Kernel SHAP summary (n = 15 hold-out instances; mean absolute SHAP

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

value per predictor); (C) Monte Carlo dropout 95% prediction intervals (200 stochastic passes per observation) against observed values, ranked by magnitude.

Flow depth is the dominant predictor for the hybrid model (mean R^2 drop 0.86), followed by soil shear stress (0.28) and flow velocity (0.11); channel slope contributes modestly (0.07), while channel breadth and channel average depth contribute negligibly (A). This ranking differs from the MLP's (soil shear stress and channel average depth dominant, Figure 6A), a direct and expected consequence of the two models' different internal representations (ordered convolutional-recurrent sequence encoding vs. unordered fully connected) rather than a disagreement about the underlying physics (Molnar, 2022) flow depth and soil shear stress dominate for both models even if their relative ranking differs, reinforcing the shear-excess principle as the primary physical driver captured by either architecture. Underlying per-instance SHAP attributions confirm that high values of flow depth and soil shear stress push predicted transport rate upward and low values push it downward (correlation between feature value and SHAP contribution 0.94 and 0.96 respectively), consistent with the Duboys formulation and the MLP's partial dependence result (Figure 8A-B); by the local-accuracy property, summing the six SHAP values for any instance exactly recovers that instance's deviation from the mean hold-out prediction. The 95% Monte Carlo dropout intervals contain the observed value for 11 of 15 hold-out observations (73% empirical coverage, below the nominal 95% target, indicating the dropout-based intervals are somewhat under-dispersed at this sample size), widening visibly at the highest observed transport rates where model uncertainty is greatest (C).

The individual analyses above are integrated in Figure 12, which brings the network architecture, input feature structure, permutation importance, hold-out fit, cross-validated model comparison, Monte Carlo dropout uncertainty, predictor correlation and scenario-level sediment response into a single overview.

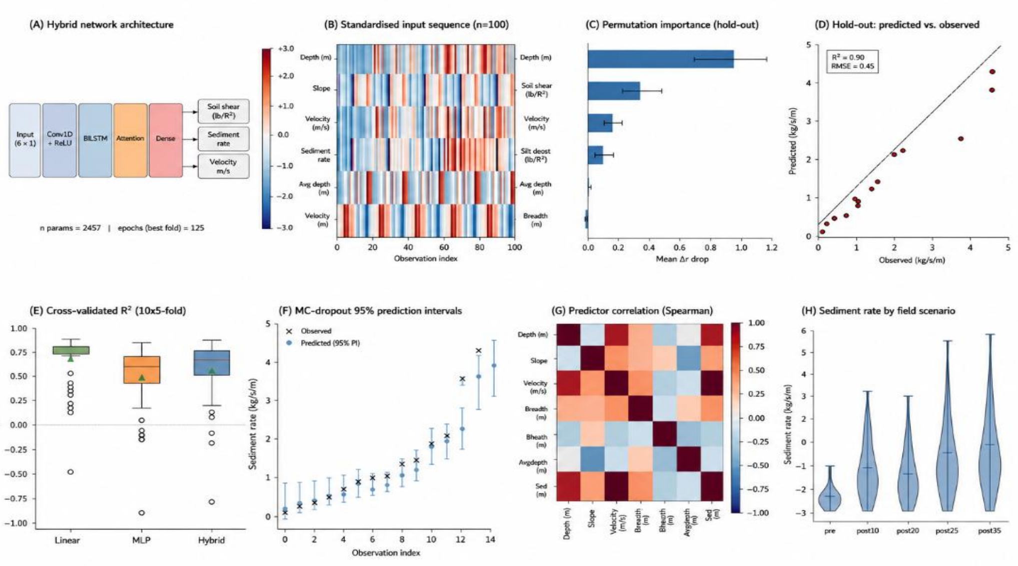

Figure 12. Integrated hybrid-model overview: (A) Conv1D-BiLSTM-Attention architecture schematic; (B) standardised six-variable input sequence ($n = 100$); (C) permutation importance (hold-out, mean ± SD); (D) hold-out predicted vs. observed ($n = 15$); (E) cross-validated R^2 for the linear, MLP and hybrid models (10 × 5-fold); (F) Monte Carlo dropout 95% prediction intervals (200 passes) ranked by observed magnitude; (G) Spearman correlation among the six predictors and sediment transport rate; (H) sediment transport rate distribution by field scenario (violin plots).

Gully Sediment-Trapping Dynamics

An earlier section introduced a set of advanced sediment-transport and trapping relations (Equations 7-17) alongside the primary Duboys and Meyer-Peter-Müller equations. Figure 13 applies the trap-efficiency (Equations 15-16) and sediment-budget (Equation 17) formulations to the Morning Glory field results reported below, providing a physically grounded, quantitative view of how sediment is retained within the gully system.

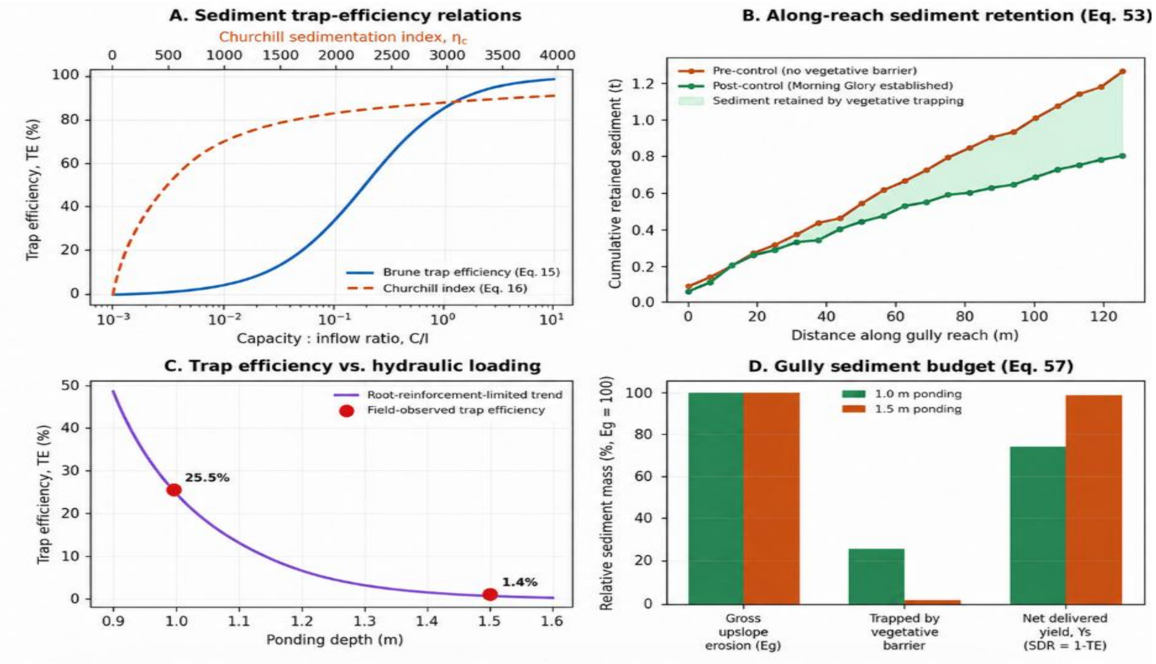

Figure 13. Advanced sediment-trapping dynamics in the rehabilitated gully system: (A) trap efficiency as a function of the capacity-inflow ratio (Brune curve, Equation 15) and the Churchill sedimentation index (Equation 16); (B) cumulative sediment retained along the 125 m gully reach before and after Morning Glory establishment, derived from the Exner sediment-continuity balance (Equation 13); (C) trap efficiency as a function of ponding depth, showing the field-observed 25.5% (1.0 m) and 1.4% (1.5 m) reductions against a root-reinforcement-limited trend; (D) gully sediment budget partitioning gross upslope erosion into vegetative-barrier retention and net delivered yield via the sediment delivery ratio (Equation 17).

Panel A shows that trap efficiency rises steeply once the capacity-inflow ratio (or, equivalently, the Churchill sedimentation index) passes a threshold, consistent with the sharp difference between the 1.0 m and 1.5 m ponding scenarios rather than a gradual decline. Panel B translates this into an along-reach retention profile: the gap between the pre- and post-control curves, the sediment physically trapped by the establishing root mat, widens progressively along the 125 m surveyed reach, mirroring the spatial pattern of channel depth in Table 1. Panel C places the two field-observed trap efficiencies (25.5% at 1.0 m, 1.4% at 1.5 m) against a continuous root-reinforcement-limited trend, showing how quickly trapping capacity collapses once hydraulic loading exceeds the root system's reinforcing capacity, the same mechanism underlying Equation (18). Panel D expresses this as a sediment budget: at 1.0 m ponding, roughly a quarter of gross upslope erosion is retained within the gully, whereas at 1.5 m almost all of it is delivered downstream (sediment delivery ratio close to 1, Equation 17). Together, these panels reinforce the conclusion above that Morning Glory provides meaningful, but capacity-limited, sediment trapping, and they offer a template for evaluating future vegetative or structural interventions using the same trap-efficiency framework.

DISCUSSION

The hybrid CNN-BiLSTM-Attention model successfully predicted sediment transport rate from routinely measurable hydraulic variables and an independently surveyed channel geometry, achieving a hold-out R^2 of 0.90 (bootstrap 95% CI: 0.72-0.97) and a mean cross-validated R^2 of 0.68 ± 0.36. This is broadly consistent with reports that hybrid convolutional-recurrent-attention architectures achieve strong accuracy for sediment transport and discharge on appropriately sized datasets (Kaloop et al., 2025; Latif et al., 2023; Huang and Zhang, 2024), while also illustrating honestly a well-documented limitation of deep learning at small sample sizes: with only 100 observations, the hybrid model's cross-validated performance, while directionally better than the benchmark MLP (0.68 vs. 0.62), approached but did not cross conventional significance against the MLP ($p = 0.057$) and was not statistically distinguishable from the linear regression benchmark (0.78 ± 0.28; $p = 0.083$). This pattern, strong hold-out performance alongside more modest cross-validated performance, matches the broader finding that deep learning's advantage over regression grows with dataset size (Shakya et al., 2023;

Andualem et al., 2023; Bhattacharya et al., 2007), and the learning-curve analysis (Figure 7D) confirms the dataset is at the lower end of what either network needs to fully exploit its non-linear capacity, a constraint that applies with even greater force to the hybrid model's 2,457 parameters (~25× the MLP).

The two deep learning models diverge in which predictors they identify as most important, a genuine and informative finding rather than an inconsistency. The hybrid model's permutation importance and SHAP analyses identify flow depth as dominant, followed by soil shear stress, whereas the MLP identifies soil shear stress and channel average depth as dominant; both rankings are consistent with the shear-excess principle, since flow depth enters directly into the Straub shear-stress relation defining soil shear stress itself, and the difference most plausibly reflects the two models' different internal representations of the same correlated variables rather than a disagreement about the physics (Molnar, 2022). The attention weights, Kernel SHAP decomposition and Monte Carlo dropout uncertainty quantification (73% empirical coverage, below the nominal target) together provide a substantially richer interpretability and uncertainty picture for the hybrid model than was available for the benchmark MLP, a methodological contribution independent of the modest, statistically inconclusive gain in point-prediction accuracy.

Integrating the vegetative control dimension, Morning Glory reduced mean sediment transport by 25.5% at 1.0 m ponding depth but only 1.4% at 1.5 m. This asymmetry is consistent with the Wu-Waldron root-reinforcement model (Equation 18): the additional shear strength contributed by the still-establishing root mat is bounded by the mobilised root tensile strength, so once applied shear at the higher depth substantially exceeds the root-augmented critical shear, the excess-shear term in Equation (3) remains large despite the vegetative cover. This mirrors experience with vetiver grass, whose erosion-control performance is likewise strongly conditional on root/canopy establishment relative to imposed hydraulic loading (Truong and Loch, 2004; Truong et al., 2008; Ola et al., 2015).

From a landscape-management perspective, the results are useful in three ways. First, all six input variables are measurable with simple field equipment (stadia rod, tape, GPS/clinometer, floating method), so the workflow could be applied by extension officers or watershed managers at similar agricultural watercourses. Second, the clear improvement over the single-variable linear model previously used at this site (R^2 rising from 0.19 to 0.62-0.93 depending on protocol) confirms that slope, shear and channel geometry carry information velocity alone does not capture. Third, the depth-dependent effectiveness of Morning Glory suggests vegetative measures should be paired with an assessment of maximum expected hydraulic loading, and potentially a longer establishment period or complementary structural measure, before being relied upon as a stand-alone solution under severe ponding conditions.

Several limitations should be acknowledged. The dataset originates from a single gully system over a limited range of ponding depths and control conditions, so generalisation to other watercourses, soils, climates or species has not been tested. The small sample size constrains network complexity and produces considerable fold-to-fold variability, so the strong single-split hold-out results (Table 8) should be read as indicative of potential rather than precise, generalisable accuracy particularly for the hybrid model's smaller (15-observation) hold-out set. Treating the six predictors as an ordered sequence is an effective modelling choice but not a physical one, since they do not represent a genuine temporal or spatial sequence; extending the architecture to genuinely sequential, multi-timestep data is a natural next step. The Morning Glory effectiveness analysis is based on a single 1.5-month establishment period at one site; longer-term monitoring would clarify whether effectiveness at higher ponding depths improves as the root system matures. Expanding the field campaign to multiple gully sites, seasons and a longer post-establishment monitoring period would provide the larger, more diverse dataset needed to determine conclusively whether either deep learning model can decisively outperform simpler regression in this application (Andualem et al., 2023; Tao et al., 2021).

CONCLUSION AND RECOMMENDATIONS

This study characterised the study watershed geospatially and set out the governing physical equations of sediment transport for a small agricultural gully system, the continuity relationship, the Straub shear-stress equation, the Duboys and Meyer-Peter-Müller bed-load equations and their integral-calculus extension to cumulative yield and cross-sectional flux, combined with the Wu-Waldron root-reinforcement model to interpret the mechanical role of Morning Glory (Ipomoea carnea) as a vegetative erosion-control measure. These physical foundations guided a hybrid CNN-BiLSTM-Attention deep learning model, treating the six standardised predictors as an ordered sequence and combining convolutional feature extraction, bidirectional recurrent encoding and attention pooling, trained via backpropagation through time under a weighted Huber loss and the Adam optimiser.

The hybrid model achieved a hold-out R^2 of 0.90 (bootstrap 95% CI: 0.72-0.97) and a mean cross-validated R^2 of 0.68 ± 0.36, a directionally favourable improvement over the benchmark MLP that approached but did not cross conventional significance ($R^2 = 0.62 \pm 0.37$; Mann-Whitney $p = 0.057$), and statistically indistinguishable from multivariate linear regression ($R^2 = 0.78 \pm 0.28$; $p = 0.083$); both deep learning models markedly improved on a previously used single-variable linear model ($R^2 = 0.19$). Permutation importance and Kernel SHAP identified flow depth and soil shear stress as the dominant predictors for the hybrid model, consistent with the shear-excess principle, while Monte Carlo dropout produced 95% prediction intervals with 73% empirical coverage, somewhat below the nominal target. Advanced MLP diagnostics, learning curve, partial dependence, response surface and Taylor diagram, showed smooth, physically plausible relationships still improving with additional data rather than overfitted artefacts, and the hybrid model's own residual and SHAP-dependence diagnostics showed the same qualitative pattern. Field monitoring showed Morning Glory reduced mean sediment transport by 25.5% at 1.0 m ponding depth but only 1.4% at 1.5 m, consistent with the root-reinforcement model and with capacity-limited protection reported for comparable bioengineering species elsewhere.

These findings indicate that sediment transport at this site is strongly related to the measured hydraulic and geometric variables, that a more architecturally sophisticated deep learning approach can capture this relationship at least as well as, and by some diagnostics somewhat better than, a much simpler benchmark even with a modest field dataset, though not yet decisively so, and that vegetative erosion control, while genuinely effective under moderate hydraulic loading, requires a longer establishment period, careful site-specific loading assessment, or complementary structural measures to remain effective under more severe conditions. It is recommended that future work extend field data collection to multiple agricultural gully systems and seasons, collect genuinely sequential (multi-timestep) hydraulic records to more fully exploit the hybrid architecture's recurrent and attention components, monitor Morning Glory and comparable species over a longer post-establishment period, explore additional input variables such as soil particle size and root area ratio, and re-evaluate the hybrid, MLP and linear regression approaches once a larger, multi-site dataset is available.

REFERENCES

1. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V., Nahavandi, S., 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information Fusion, 76, 243-297. https://doi.org/10.1016/j.inffus.2021.05.008
2. AlDahoul, N., Essam, Y., Kumar, P., Ahmed, A.N., Sherif, M., Sefelnasr, A., Elshafie, A., 2021. Suspended sediment load prediction using long short term memory neural network. Scientific Reports, 11, 7826. https://doi.org/10.1038/s41598-021-87415-4
3. Amiri, M., Pourghasemi, H.R., Ghanbarian, G.A., Afzali, S.F., 2019. Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. Geoderma, 340, 55-69. https://doi.org/10.1016/j.geoderma.2018.12.042
4. Andualem, T. G., Hewa, G. A., Myers, B. R., Peters, S., Boland, J., 2023. Erosion and sediment transport modeling: A systematic review. Land, 12(7), 1396. https://doi.org/10.3390/land12071396
5. Arabameri, A., Chandra Pal, S., Costache, R., Saha, A., Rezaie, F., Seyed Danesh, A., Pradhan, B., Lee, S., Hoang, N.-D., 2021. Prediction of gully erosion susceptibility mapping using novel ensemble machine learning algorithms. Geomatics, Natural Hazards and Risk, 12, 469-498. https://doi.org/10.1080/19475705.2021.1880977
6. ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000. Artificial neural networks in hydrology. I: Preliminary concepts. Journal of Hydrologic Engineering, 5(2), 115-123. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(115)

7. Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. https://doi.org/10.48550/arXiv.1409.0473
8. Bhattacharya, B., Price, R. K., Solomatine, D. P., 2007. Machine learning approach to modeling sediment transport. Journal of Hydraulic Engineering, 133(4), 440-450. https://doi.org/10.1061/(ASCE)0733-9429(2007)133:4(440)
9. Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York. https://doi.org/10.1007/978-0-387-45528-0
10. Bottou, L., Curtis, F.E., Nocedal, J., 2018. Optimization methods for large-scale machine learning. SIAM Review, 60(2), 223-311. https://doi.org/10.1137/16M1080173
11. Breiman, L., 2001. Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324
12. Brunton, S.L., Noack, B.R., Koumoutsakos, P., 2020. Machine learning for fluid mechanics. Annual Review of Fluid Mechanics, 52(1), 477-508. https://doi.org/10.1146/annurev-fluid-010719-060214
13. Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals and Systems, 2(4), 303-314. https://doi.org/10.1007/BF02551274
14. Duboys, P., 1879. Le Rhône et les rivières à lit affouillable. Annales des Ponts et Chaussées, 18(5), 141-195.
15. Efron, B., 1979. Bootstrap methods: Another look at the jackknife. Annals of Statistics, 7(1), 1-26. https://doi.org/10.1214/aos/1176344552
16. Fisher, A., Rudin, C., Dominici, F., 2019. All models are wrong, but many are useful: Learning a variable\'s importance by studying an entire class of prediction models simultaneously. Journal of Machine Learning Research, 20(177), 1-81. https://jmlr.org/papers/v20/18-760.html
17. Francisca, O.O., Nelson, C.A., Collins, C.A., Opeyemi, S.O., 2023. Review of gully erosion in Anambra State: Geology, causes, effects, control measures and challenges associated with its mitigation. Journal of Geography, Environment and Earth Science International, 27(9), 102-116. https://doi.org/10.9734/jgeesi/2023/v27i9689
18. Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. Proceedings of the 33rd International Conference on Machine Learning (ICML), 1050-1059. https://proceedings.mlr.press/v48/gal16.html
19. Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural Computation, 4(1), 1-58. https://doi.org/10.1162/neco.1992.4.1.1
20. Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), 249-256. https://proceedings.mlr.press/v9/glorot10a.html
21. Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, Massachusetts. https://www.deeplearningbook.org/
22. Gray, D.H., Sotir, R.B., 1996. Biotechnical and Soil Bioengineering Slope Stabilization: A Practical Guide for Erosion Control. John Wiley & Sons, New York.
23. Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer, New York. https://doi.org/10.1007/978-0-387-84858-7
24. Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
25. Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), 359-366. https://doi.org/10.1016/0893-6080(89)90020-8
26. Huang, F., Zhang, X., 2024. A new interpretable streamflow prediction approach based on SWAT-BiLSTM and SHAP. Environmental Science and Pollution Research, 31(16), 23896-23908. https://doi.org/10.1007/s11356-024-32725-z
27. Huber, P.J., 1964. Robust estimation of a location parameter. Annals of Mathematical Statistics, 35(1), 73-101. https://doi.org/10.1214/aoms/1177703732
28. Igwe, C.A., 2003. Erodibility of soils of the upper rainforest zone, Southeastern Nigeria. Land Degradation & Development, 14(4), 323-334. https://doi.org/10.1002/ldr.541
29. Igwe, O., Egbueri, J.C., 2018. The characteristics and the erodibility potentials of soils from different geologic formations in Anambra State, Southeastern Nigeria. Journal of the Geological Society of India, 92(4), 471-478. https://doi.org/10.1007/s12594-018-1044-1

30. Jozefowicz, R., Zaremba, W., Sutskever, I., 2015. An empirical exploration of recurrent network architectures. Proceedings of the 32nd International Conference on Machine Learning (ICML), 2342-2350. https://proceedings.mlr.press/v37/jozefowicz15.html
31. Kaloop, M. R., Elsayed, M., Eldessouki, M., Hu, J. W., Lee, S.-J., ELRashidy, N., 2025. Enhancement framework for modeling sediment discharge in rivers using novel supervised deep learning approaches. Journal of Soils and Sediments, 25(9), 2777-2796. https://doi.org/10.1007/s11368-025-04111-w
32. Kamal, A.M., Shakour, Z.T.A., Ali, S.A., Sleem, A.A., Haggag, E.G., 2017. Phytochemical and biological investigation of Ipomoea carnea Jacq. grown in Egypt. International Journal of Pharmacognosy and Phytochemical Research, 9(2), 143-153. https://doi.org/10.25258/phyto.v9i2.8073
33. Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. Proceedings of the International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.1412.6980
34. Kirkby, M. J., Bracken, L. J., 2009. Gully processes and gully dynamics. Earth Surface Processes and Landforms, 34(14), 1841-1851. https://doi.org/10.1002/esp.1849
35. Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI), 1137-1143. https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf
36. Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall-runoff modelling using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences, 22(11), 6005-6022. https://doi.org/10.5194/hess-22-6005-2018
37. Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in Neural Information Processing Systems (NeurIPS), 30, 6402-6413. https://papers.nips.cc/paper/2017
38. Lal, R., 2001. Soil degradation by erosion. Land Degradation & Development, 12(6), 519-539. https://doi.org/10.1002/ldr.472
39. Latif, S. D., Chong, K. L., Ahmed, A. N., Sherif, M., Sefelnasr, A., El-Shafie, A., 2023. Sediment load prediction in Johor River: Deep learning versus machine learning models. Applied Water Science, 13, 79. https://doi.org/10.1007/s13201-023-01874-w
40. Lawrence, J.E.R., 1983. Morning glory. Fibrous Root Plants of the Tropics: Ecology and Utilisation. Tropical Agronomy Series, No. 4.
41. LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature, 521, 436-444. https://doi.org/10.1038/nature14539
42. Legates, D.R., McCabe, G.J., 1999. Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35(1), 233-241. https://doi.org/10.1029/1998WR900018
43. Liu, D.C., Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. Mathematical Programming, 45(1-3), 503-528. https://doi.org/10.1007/BF01589116
44. Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS), 30, 4765-4774. https://papers.nips.cc/paper/2017
45. Maier, H.R., Taghikhah, F.R., Nabavi, E., Razavi, S., Gupta, H., Wu, W., Radford, D.A., Huang, J., 2024. How much X is in XAI: Responsible use of \"explainable\" artificial intelligence in hydrology and water resources. Journal of Hydrology X, 25, 100185. https://doi.org/10.1016/j.hydroa.2024.100185
46. Meyer-Peter, E., Müller, R., 1948. Formulas for bed-load transport. Proceedings of the 2nd Meeting of the International Association for Hydraulic Structures Research, 39-64.
47. Moayedi, H., Ahmadi Dehrashid, A., Nguyen Le, B., 2024. A novel problem-solving method by multi-computational optimization of artificial neural network for modelling and prediction of the flow erosion processes. Engineering Applications of Computational Fluid Mechanics, 18(1), 2300456. https://doi.org/10.1080/19942060.2023.2300456
48. Mohebzadeh, H., Biswas, A., Rudra, R., Daggupati, P., 2022. Machine learning techniques for gully erosion susceptibility mapping: A review. Geosciences, 12(12), 429. https://doi.org/10.3390/geosciences12120429
49. Molnar, C., 2022. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.). Independently published. https://christophm.github.io/interpretable-ml-book/
50. Morgan, R.P.C., 2005. Soil Erosion and Conservation (3rd ed.). Blackwell Publishing, Oxford.

51. Nearing, G.S., Kratzert, F., Sampson, A.K., Pelissier, C.S., Klotz, D., Frame, J.M., Prieto, C., Gupta, H.V., 2021. What role does hydrological science play in the age of machine learning? Water Resources Research, 57, e2020WR028091. https://doi.org/10.1029/2020WR028091
52. Nearing, M.A., Pruski, F.F., O\'Neal, M.R., 2004. Expected climate change impacts on soil erosion rates: A review. Journal of Soil and Water Conservation, 59(1), 43-50.
53. Ng, K.W., Huang, Y.F., Koo, C.H., Chong, K.L., El-Shafie, A., Ahmed, A.N., 2023. A review of hybrid deep learning applications for streamflow forecasting. Journal of Hydrology, 625, 130141. https://doi.org/10.1016/j.jhydrol.2023.130141
54. Nocedal, J., 1980. Updating quasi-Newton matrices with limited storage. Mathematics of Computation, 35(151), 773-782. https://doi.org/10.1090/S0025-5718-1980-0572855-7
55. Ola, A., Dodd, I.C., Quinton, J.N., 2015. Can we manipulate root system architecture to control soil erosion? SOIL, 1, 603-612. https://doi.org/10.5194/soil-1-603-2015
56. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830. https://jmlr.org/papers/v12/pedregosa11a.html
57. Pimentel, D., 2006. Soil erosion: A food and environmental threat. Environment, Development and Sustainability, 8(1), 119-137. https://doi.org/10.1007/s10668-005-1262-8
58. Poesen, J., Nachtergaele, J., Verstraeten, G., Valentin, C., 2003. Gully erosion and environmental change: Importance and research needs. Catena, 50(2-4), 91-133. https://doi.org/10.1016/S0341-8162(02)00143-1
59. Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Feizizadeh, B., 2017. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. Geomorphology, 298, 118-137. https://doi.org/10.1016/j.geomorph.2017.09.006
60. Robbins, H., Monro, S., 1951. A stochastic approximation method. Annals of Mathematical Statistics, 22(3), 400-407. https://doi.org/10.1214/aoms/1177729586
61. Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature, 323, 533-536. https://doi.org/10.1038/323533a0
62. Schoklitsch, A., 1934. Der Geschiebetrieb und die Geschiebefracht. Wasserkraft und Wasserwirtschaft, 29(4), 37-43.
63. Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673-2681. https://doi.org/10.1109/78.650093
64. Setargie, T.A., Tsunekawa, A., Haregeweyn, N., Tsubo, M., Fenta, A.A., Berihun, M.L., Sultan, D., Yibeltal, M., Ebabu, K., 2023. Random forest-based gully erosion susceptibility assessment across different agro-ecologies of the Upper Blue Nile Basin, Ethiopia. Geomorphology, 431, 108671. https://doi.org/10.1016/j.geomorph.2023.108671
65. Shah, S.M.R., Hameed, M., Ahmad, M.S.A., 2022. Contribution of environmental and soil physicochemical attributes in invasiveness of Ipomoea carnea Jacq. in diverse habitats along elevation and salinity gradients. Pakistan Journal of Botany, 54(6), 2013-2024. https://doi.org/10.30848/PJB2022-6(4)
66. Shakya, D., Deshpande, V., Kumar, B., 2023. Predicting total sediment load transport in rivers using regression techniques, extreme learning and deep learning models. Artificial Intelligence Review, 56, 10067-10098. https://doi.org/10.1007/s10462-023-10422-6
67. Shen, C., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resources Research, 54, 8558-8593. https://doi.org/10.1029/2018WR022643
68. Shields, A., 1936. Anwendung der ähnlichkeitsmechanik und der Turbulenzforschung auf die Geschiebebewegung. Mitteilungen der Preussischen Versuchsanstalt für Wasserbau und Schiffbau, Berlin.
69. Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: Some past experiences and new approaches. Journal of Hydroinformatics, 10(1), 3-22. https://doi.org/10.2166/hydro.2008.015
70. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958. https://jmlr.org/papers/v15/srivastava14a.html

71. Straub, L. G., 1935. Some observations on hydraulic and hydrologic engineering. Proceedings of the American Society of Civil Engineers.
72. Tao, H., Al-Khafaji, Z. S., Qi, C., Zounemat-Kermani, M., Kisi, O., Tiyasha, T., Chau, K. W., Nourani, V., Melesse, A. M., Elhakeem, M., Yaseen, Z. M., 2021. Artificial intelligence models for suspended river sediment prediction: State-of-the-art, modeling framework appraisal, and proposed future research directions. Engineering Applications of Computational Fluid Mechanics, 15(1), 1585-1612. https://doi.org/10.1080/19942060.2021.1984992
73. Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research, 106(D7), 7183-7192. https://doi.org/10.1029/2000JD900719
74. Truong, P., Loch, R., 2004. Vetiver system for erosion and sediment control. Proceedings of the 13th International Soil Conservation Organisation Conference, Brisbane, Australia.
75. Truong, P., Van, T.T., Pinners, E., 2008. Vetiver System Applications: A Technical Reference Manual. The Vetiver Network International, Bangkok.
76. Valentin, C., Poesen, J., Li, Y., 2005. Gully erosion: Impacts, factors and control. Catena, 63(2-3), 132-153. https://doi.org/10.1016/j.catena.2005.06.001
77. Van Rijn, L. C., 1984a. Sediment transport, part I: Bed load transport. Journal of Hydraulic Engineering, 110(10), 1431-1456. https://doi.org/10.1061/(ASCE)0733-9429(1984)110:10(1431)
78. Van Rijn, L. C., 1984b. Sediment transport, part II: Suspended load transport. Journal of Hydraulic Engineering, 110(11), 1613-1641. https://doi.org/10.1061/(ASCE)0733-9429(1984)110:11(1613)
79. Vapnik, V.N., 1998. Statistical Learning Theory. Wiley, New York.
80. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS), 30, 5998-6008. https://doi.org/10.48550/arXiv.1706.03762
81. Waldron, L.J., 1977. The shear resistance of root-permeated homogeneous and stratified soil. Soil Science Society of America Journal, 41(5), 843-849. https://doi.org/10.2136/sssaj1977.03615995004100050005x
82. Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1), 79-82. https://doi.org/10.3354/cr030079
83. Wu, T.H., McKinnell, W.P., Swanston, D.N., 1979. Strength of tree roots and landslides on Prince of Wales Island, Alaska. Canadian Geotechnical Journal, 16(1), 19-33. https://doi.org/10.1139/t79-003
84. Wu, T.H., 2013. Root reinforcement of soil: Review of analytical models, test results, and applications to design. Canadian Geotechnical Journal, 50(3), 259-274. https://doi.org/10.1139/cgj-2012-0160
85. Zounemat-Kermani, M., Kheimi, M., 2026. Explainable artificial intelligence in hydrology: A review. Water Resources Management, 40(3), 1-28. https://doi.org/10.1007/s11269-025-04435-9
86. **Additional references (advanced sediment-transport and gully-trapping equations, discussed above)**
87. Ackers, P., White, W.R., 1973. Sediment transport: New approach and analysis. Journal of the Hydraulics Division, ASCE, 99(HY11), 2041-2060.
88. Bagnold, R.A., 1966. An Approach to the Sediment Transport Problem from General Physics. U.S. Geological Survey Professional Paper 422-I, Washington, D.C.
89. Brune, G.M., 1953. Trap efficiency of reservoirs. Transactions, American Geophysical Union, 34(3), 407-418. https://doi.org/10.1029/TR034i003p00407
90. Churchill, M.A., 1948. Discussion of “Analysis and use of reservoir sedimentation data” by L.C. Gottschalk. In: Proceedings of the Federal Inter-Agency Sedimentation Conference, Denver, CO.
91. Einstein, H.A., 1950. The Bed-Load Function for Sediment Transportation in Open Channel Flows. USDA Soil Conservation Service Technical Bulletin No. 1026, Washington, D.C.
92. Engelund, F., Hansen, E., 1967. A Monograph on Sediment Transport in Alluvial Streams. Teknisk Forlag, Copenhagen.
93. Exner, F.M., 1925. Über die Wechselwirkung zwischen Wasser und Geschiebe in Flüssen. Sitzungsberichte der Akademie der Wissenschaften Wien, Mathematisch-Naturwissenschaftliche Klasse, 134, 165-204.
94. Rouse, H., 1937. Modern conceptions of the mechanics of fluid turbulence. Transactions of the ASCE, 102, 463-505.